UNIVERSITY OF TARTU
Institute of Computer Science
Software Engineering Curriculum

Anette Maria Kuklane

Data Mesh for Financial Service Providers

Master's Thesis (30 ECTS)

Supervisor: Fredrik Milani, PhD

Tartu 2024

Data Mesh for Financial Service Providers

Abstract:
The volume of data and its generation rate are increasing at unprecedented levels each year. Most organizations depend on centralized architectures for analytical data, which struggle to cope with the growing demands for data. In response, data mesh has emerged as a new architectural approach that decentralizes data management into domains to address the limitations of centralized data architectures. Due to the novelty of data mesh, there is a lack of research on its necessity and applicability in the financial services sector. This thesis sought to explore the current data management practices of financial service providers and how these practices might limit the achievement of business needs and objectives. Moreover, it aimed to investigate how data mesh can be implemented within the financial services sector, focusing on defining data domains and exploring the value data mesh could provide. For this, semi-structured interviews were conducted with ten industry specialists and analyzed using the thematic analysis method. The study found that the impact of data management practices on business objectives depends on how long they have been in place and the variety of products offered. Traditional banks faced more limitations compared to challenger banks and FinTechs. Similarly, the adoption of a data mesh architecture could be particularly valuable for larger and more established institutions, such as traditional banks. However, while other financial organizations may see less immediate value, early adoption could be beneficial for future scaling. Additionally, the findings indicated that adopting even some of the data mesh principles, such as decentralized data ownership and governance, could benefit financial organizations. Although there seems to be no clear way of how to define the data domains, the findings suggested a product or process-oriented approach that aligns with business domains as well as highlighted that certain data sets should be centralized.

Keywords:
Data mesh, data architecture, financial service providers, financial services sector

CERCS:
P170 Computer science, numerical analysis, systems, control

Andmevõrk finantsteenuste pakkujate jaoks

Lühikokkuvõte:
Andmemahud ja uute andmete genereerimise kiirus kasvab igal aastal ennenägematul tasemel. Enamik organisatsioone kasutab analüütiliste andmete jaoks tsentraliseeritud arhitektuure, mis ei pruugi toime tulla kasvava nõudlusega. Selle probleemi lahendamiseks on tekkinud uus arhitektuuriline lähenemine, milleks on andmevõrk (ingl *data mesh*). Andmevõrk aitab lahendada tsentraliseeritud andmearhitektuuride poolt tekkinud piiranguid detsentraliseerides andmete halduse erinevate domeenide vahel. Kontseptsiooni uudsuse tõttu puuduvad uurimustööd andmevõrgu vajalikkuse ja rakendamise kohta finantsteenuste sektoris. Käesoleva magistritöö eesmärgiks oli uurida, kuidas finantsteenuste pakkujad andmeid haldavad ja kuidas kasutuses olevad praktikad võivad takistada ärivajaduste ja eesmärkide täitmist. Lisaks oli töö eesmärgiks uurida, kuidas andmevõrgu arhitektuuri oleks võimalik finantsteenuste sektoris rakendada, keskendudes seejuures domeenide defineerimisele ja andmevõrgust saadava kasu uurimisele. Selleks viidi läbi poolstruktureeritud intervjuud kümne valdkonnaspetsialistiga, misjärel analüüsiti intervjuusid temaatilise analüüsi meetodil. Tulemused näitasid, et andmehalduspraktikate mõju ärieesmärkidele sõltub sellest, kui kaua on need kasutusel olnud ning samuti tootevaliku mitmekesisusest. Seejuures traditsioonilised pangad seisavad silmitsi rohkemate takistustega võrreldes väljakutsuvate pankade (ingl *challenger bank*) ja FinTech ettevõtetega. Sarnaselt eelnevale selgus, et andmevõrgu arhitektuuri kasutuselevõtt on suurema väärtusega traditsioonilistele pankadele. Teistel finantsorganisatsioonidel ei pruugi kasu andmevõrgu arhitektuurist olla kohene, kuid selle varajane rakendamine võib osutuda kasulikuks organisatsiooni kasvamisel. Lisaks näitasid tulemused, et finantsorganisatsioonidel võib kasu olla ka vaid mõne andmevõrgu põhimõtte, nagu detsentraliseeritud andmehalduse ja -valitsemise, rakendamisest. Kuigi töös ei selgunud üht kindlat andmedomeenide määratlemise viisi, viitasid tulemused toote- või protsessipõhistele domeenidele, mis peaksid olema kooskõlas ärivaldkondadega. Siiski selgus tulemustest, et teatud andmeid tuleks hoida tsentraliseeritud kujul.

Võtmesõnad:
Andmevõrk, andmearhitektuur, finantsteenuste pakkujad, finantsteenuste sektor

CERCS:
P170 Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine (automaatjuhtimisteooria)

Terms and Notations

Business domain is a segment that focuses on a specific area or business activity [1]. It's often aligned with the organization's structure, such as operations, sales, and marketing.

Challenger bank is a licensed financial institution aiming to compete with larger, established banks [2]. This thesis considers a challenger bank to be a smaller or medium-sized bank with limited or no physical presence that targets a specific niche or service, setting itself apart through focused specialization and distinctive offerings.

Data domain refers to an organization-specific grouping of related data, where the domains have ownership and control over the data they produce. This thesis uses the term synonymously with the term "domain."

Data governance refers to managing an organization's data assets through established processes, policies, and standards to ensure data availability, quality, security, integrity, and compliance [3].

Data lake is a storage repository that holds raw data, either structured, semi-structured, or unstructured, in its native format [4]. It allows for storing data from multiple sources, enabling comprehensive analysis of the stored data [4].

Data management refers to the practices and strategies for managing data in an organization.

Data management practices are the methods and procedures employed in managing data throughout its lifecycle, including its collection, storage, access, and organization.

Data mesh is an architectural approach that decentralizes data management [5]. It assigns ownership and responsibility for data to domain-specific teams, promotes a self-serve platform for data access, and federated computational governance to ensure organizational consistency and compliance [5].

Data mesh architecture refers to the architectural design that supports implementing the data mesh principles.

Data product is a logical entity designed to process and store data for analytics use cases, encapsulating necessary components such as input and output ports, documentation, data storage, and ownership [6].

Data warehouse is a storage repository designed to store, manage, and retrieve structured data from various sources [7]. It is optimized for historical and query-intensive data, supporting business intelligence (BI) activities such as analytics and structured or ad hoc queries [7].

Domain-driven design (DDD) is a software development approach that focuses on creating software models based on the reality of business domains and their related logic and behavior [8].

ETL pipeline (short for extract, transform, load pipeline), or data pipeline, is a set of data processing steps that involve extracting the data from the system(s), cleaning and transforming the data into a suitable format, and loading it into the systems for storage, data analytics and other business purposes [9].

Financial goods are tangible (such as cash or gold bars) or intangible financial items (such as stocks, insurance policies, or bank accounts) with monetary value that can be bought, sold, or traded in the financial market [10].

Financial service provider is a financial entity, a banking or a non-banking financial institution, that offers financial services (such as investment services or money transmission).

Financial services are the actions and activities related to managing financial goods, including but not limited to investment advice and loan processing [10].

Financial services sector is a part of the economy comprising companies and institutions offering financial services to businesses and individual consumers [10]. This thesis uses the term synonymously with the terms “financial sector” and “financial industry.”

FinTech company (short for a financial technology company) is a business that leverages innovative technology to create, enhance, and automate financial services and processes [11]. This thesis uses the term synonymously with the term “FinTech.”

Non-banking financial institution (NBFI) is a financial entity that offers services like investment, insurance, and money transfers without holding a traditional banking license and the associated regulations [12].

Thematic analysis is a qualitative data analysis method for finding, analyzing, and reporting patterns within data [13].

Traditional bank is a financial institution that offers in-person services at physical locations and online banking solutions, is licensed to accept deposits, provide loans, and offers various other financial services under strict regulatory standards [14].

Table of Contents

List of Figures

List of Tables

1 Introduction

The volume of data and the speed at which it is being generated is increasing at an unprecedented rate each year [15]. Data is already transforming how businesses operate and will become an increasingly critical asset for organizations [15]. Most organizations rely on centralized data architectures. These architectures, however, often fail to meet the escalating demands for data, leading to the central data team becoming a bottleneck [16], [17], causing unclear responsibilities between the domain experts and the central data team [16], and data silos as data becomes isolated in separate parts of an organization [16], [18]. Consequently, this hinders the effective utilization of data.

Data mesh has emerged in response to these challenges as a new architectural approach that decentralizes data ownership into domains [5]. Integral to the data mesh concept, data domains are the organization-specific groupings of related data, where the domains have ownership and control over the data they produce. As the data mesh concept is still relatively novel [16], [17], [19], it is viewed as being challenging to put into practice and too broad to be adopted out of the box [16]. Hence, there is a lack of insights into defining the data domains that determine what data is accessed and establish clear data ownership. Experts, however, have recognized the significance of clearly defined domains as vital for maintaining proper oversight and facilitating efficient collaboration across different domains [20].

Adopting a data mesh architecture could revolutionize data management for companies across all sectors, enabling businesses to make agile and data-driven decisions to maintain a competitive edge. The financial services sector, greatly influenced by data, stands out for its rigorous regulatory environment and diverse technological maturity, ranging from traditional banks to newer financial technology companies. Given the diversity in business domains, needs, and objectives across various sectors, this research specifically targets the financial services sector.

The research objective of this thesis is to explore how the data mesh domains can be defined for financial service providers. It aims to investigate the current data management practices and their impact on achieving business needs and objectives as well as to explore the value of adopting the data mesh architecture for financial service providers. Overall, the thesis aims to answer the following research questions (RQs):

RQ1. How is data currently managed by financial service providers?

RQ2. How do the current data management practices limit financial service providers in achieving their business needs and objectives?

RQ3. How can the data domains be defined for financial service providers to facilitate achieving business needs and objectives?

RQ4. How valuable could the adoption of a data mesh architecture be for financial service providers?

This thesis has two contributions. The first contribution is an in-depth understanding of how data is managed by financial service providers, including the current practices and how these

practices may impede meeting business needs and objectives. The second contribution is a deeper knowledge of how data mesh can be defined within the financial services sector, which involves defining the data domains and assessing the value data mesh could bring.

The contributions of this thesis will benefit financial service providers as they provide expert insights into leveraging data mesh. Data specialists from the financial services sector can use the results of this thesis to understand whether moving towards a data mesh could be valuable and, if so, how the domains in a data mesh can be defined to start transitioning towards adopting a data mesh architecture.

To achieve this, industry specialists from different financial institutions will be interviewed. These insights are essential for getting an in-depth understanding of the subject matter to understand the specific needs of the financial service providers. Lastly, the interview results will be analyzed using the thematic analysis method, and a discussion of the results will be provided.

The following describes the structure of the thesis. Section 2 provides an overview of data mesh, data domains, and financial service providers. Section 3 discusses the related work. Section 4 gives an overview of the methodology used, specifically outlining the research questions as well as the data collection and analysis methods. Results are presented in section 5. A discussion of the results is given in section 6. Finally, the thesis is concluded in section 7.

2 Background

This section introduces the fundamental concepts used in the thesis, such as data mesh, data domains, and financial service providers.

2.1 Data Mesh

Given the vast volumes and increasing importance of data, organizations are striving to leverage advanced data analytics for more informed decision-making [16]. Data can be divided into operational and analytical data [5]. Operational data is defined as the data that is dynamically managed in real-time and stored in databases of microservices, applications, or other systems [5]. This operational data, which is used to run the business and serve the users, is consequently transformed into analytical data using data integration mechanisms such as ETL (extract, transform, load) pipelines [5]. These pipelines involve a sequence of operations that extract data from one or more operational systems, cleanse and transform it, and finally load it into a designated storage system, such as data warehouses or lakes, for subsequent analysis and reporting [9].

The concept of data mesh has emerged as existing centralized data architectures have shown their limitations regarding scalability, accessibility, and cost [17]. Data mesh applies the principles of microservices in software engineering to data engineering [17] to manage analytical data, from now on referred to as data for simplicity, at scale. It introduces four core principles: domain-oriented decentralized data ownership, data as a product, self-serve data infrastructure as a platform, and federated computational governance.

Domain-oriented decentralized ownership of data, as the first principle, is the core idea behind the data mesh concept in which the data responsibility is distributed to different domains based on their familiarity and control over the data [5], [6], [16]. This promotes engagement among domain experts, who are best equipped to understand and handle their data's specificities, resulting in better quality and more relevant data products. Teams can make quicker, more informed decisions by having immediate access to their domain's data without cross-departmental dependencies [6]. This transformation to domain-oriented data aims to enable scalable data sharing aligned with organizational growth and optimizes for continuous change within the domains [5].

The second principle of data mesh, data as a product, means that data should be treated the same way a software product would [6], [20]. Hence, data products adhere to the following characteristics to be considered useful: discoverable, addressable, understandable, trustworthy, accessible, interoperable, valuable, and secure [6], [19]. This includes having input ports to gather data from operational systems or other data products and output ports that facilitate data distribution to other domains or services [6]. The output port(s) should serve data sets as defined by a published data contract that specifies, for example, the structure and semantics of the data attributes, the quality attributes, and the conditions of data usage for the data consumers [6]. Moreover, a data product contains code that transforms raw data into a refined and usable state, documentation that clarifies its use and governance, and continuous monitoring to ensure data quality and availability [6].

Third, the principle of a self-serve data platform emphasizes a specialized infrastructure that provides domain teams with high-level, abstracted tools and services [5], [6], [16]. The platform serves as the technical foundation to perform tasks such as ingesting, transforming, and querying data without the constant intervention of the central IT department [6]. This enables the domain teams to efficiently create and manage high-quality data products and autonomously perform analytics on their domain and cross-domain data [5].

Lastly, the fourth principle is federated computational governance. This principle involves establishing standard practices for data security, privacy, and regulatory compliance [6] at a central level, with the execution and adaptation of these protocols entrusted to domain-specific teams [21]. This allows each team the flexibility and resources to apply these standards to best suit the domain's unique context and requirements [21]. For instance, this includes having interoperability policies to ensure that external domain teams use data products uniformly and documentation to explore and understand the data products [6]. This aims to create a unified data ecosystem that is compliant with regulations [6] and instill trust and understanding of data across the organization [18].

The implementation of these principles constitutes the data mesh architecture. This is illustrated in Figure 1, where each domain is responsible for creating and managing its data products. Data contracts, adapted from overarching governance policies, guide the publishing of data products, ensuring that data sharing and usage are uniform across domains. This is supported by a self-serve platform that, for example, provides a query engine for data retrieval and analysis as well as a data product catalog to discover and understand the available data products. This architecture thereby enables data democratization [16], [20], allowing organizations to better leverage their data for informed, data-driven decisions [16].

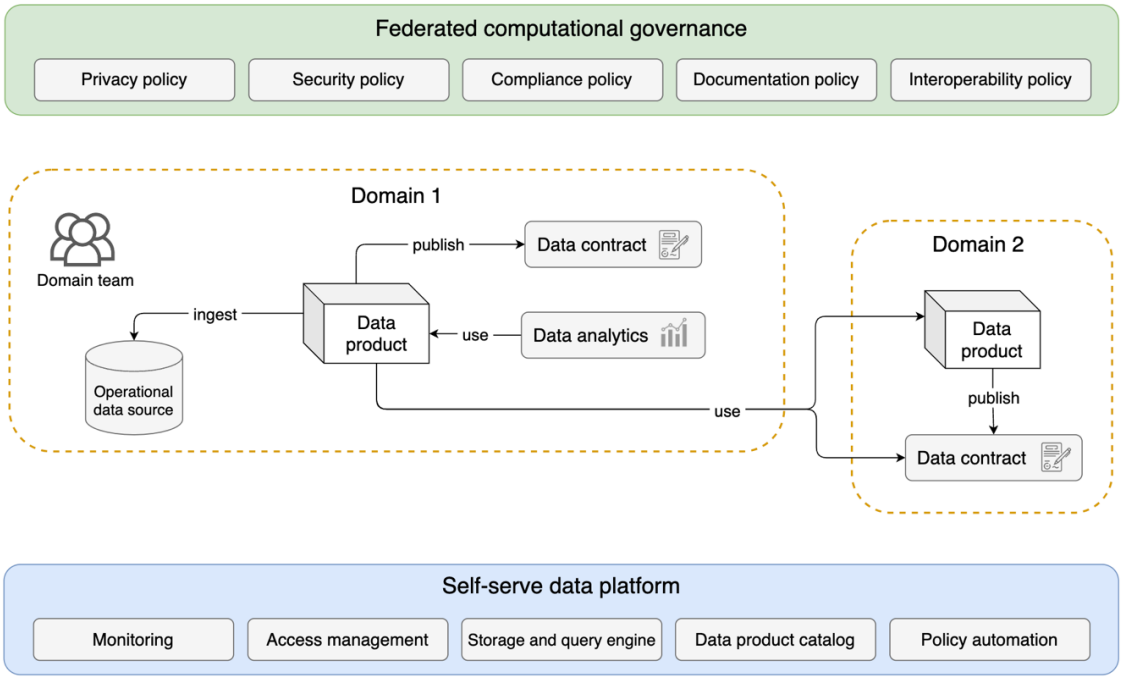

Figure 1. Overview of the data mesh architecture (adapted from [6]).

Notably, the principles within the data mesh model operate jointly to address the challenges presented by each preceding principle [22], as shown in Figure 2. Domain-oriented ownership decentralizes data control, fostering engagement but potentially leading to data silos [22]. To prevent this, data as a product ensures data remains accessible and user-friendly across domains [22]. The self-serve data platform provides tools for autonomous data handling, which could risk governance consistency; hence, federated computational governance establishes cross-domain standards, safeguarding data quality and compliance within the decentralized framework [22]. Each principle is thus a response to the challenges posed by its predecessor, creating a balanced and integrated data ecosystem.

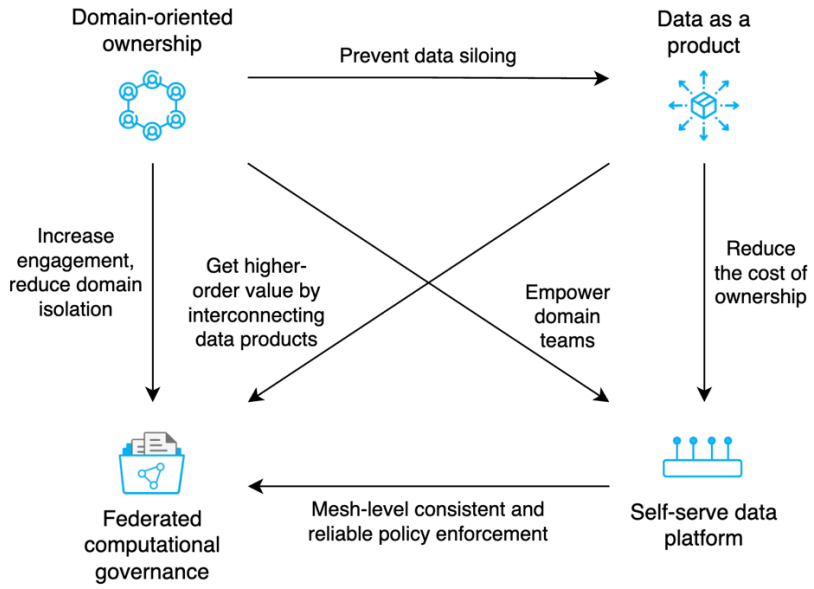

Figure 2. Interconnected principles of data mesh (adapted from [22]).

While the data mesh concept is gaining traction, it's still relatively novel and lacks extensive empirical research [16], [18], [19], [20], especially in different industry contexts. Recent studies [16], [17], [18], [19] have identified key challenges in adopting data mesh, such as transitioning to federated data governance, shifting responsibility for data product development, and managing the data product model. These challenges include concerns about security, standardization, regulatory compliance, and ensuring adequate resources and acceptance within organizations.

Despite these challenges, the potential benefits of data mesh are notable. Transitioning to this model can lead to improvements in data accessibility, increased speed in data access and product time to market, enhanced data quality, and reductions in data redundancies [16]. These changes could lead to more data-driven organizations with increased reliability in data usage.

2.2 Data Domains

Data domains could be defined based on different criteria, such as products, processes, services, or other criteria, such as market segments. For instance, organizing domains by processes involves organizing data based on operational workflows, such as account opening, transaction processing, and loan approval process. Another option is to structure

the domains by products, which organizes data according to the various financial products such as savings accounts, personal loans, and credit cards. By defining the data domains, organizations can effectively manage and analyze data according to their business needs and objectives. This ensures that data is structured and utilized to support informed decision-making and drive business value.

This approach aligns with domain-driven design (DDD), which emerged in response to the increasing complexity of software design caused by digitalization [5]. It is an approach that breaks down software design and team structuring according to the organizational structure of a business [5]. Introduced by Evans [8], DDD revolutionized modern software architecture and modeling by decomposing systems into manageable domains, exemplified by the adoption of microservice architecture [5]. However, this decentralization of operational systems has created a misalignment with how analytical data responsibilities are divided in modern organizations [5].

A core concept in DDD is the notion of bounded contexts [8]. Bounded contexts establish boundaries within which terms and concepts hold specific meanings, enabling effective communication and modeling [8]. It facilitates collaboration between domain experts and developers by ensuring a shared understanding of the domain's intricacies and requirements [8]. In the data mesh domains, bounded contexts are treated as individual data products that consist of the data, its models, and its ownership [5]. A single data domain can have multiple data products that can service different use cases [23]. Whether the domains should be defined as per the product, process, or some other criteria can depend on the specific needs and goals of the organization and the nature of the business. Hence, DDD lies in the core of the data mesh concept, prioritizing clear boundaries and domain-oriented decentralization of data, shifting the responsibility for data from the central data team to the domain teams.

In a data mesh, there are three primary archetypes of data domains [5]. The first, *source-aligned data domains*, capture the data directly from operational systems, representing "facts and reality of the business" [6:12]. This raw data is foundational, providing a truthful and immediate representation of domain activities without transformations [5]. The second archetype, *aggregate data domains*, takes a macro perspective by synthesizing information from various source-aligned domains to construct a broader, more integrated picture of business concepts [5]. This is for use cases needing a consolidated data view, such as regulatory reporting use cases. Lastly, the third, *consumer-aligned data domains*, hold the data in its most tailored form, as data is specifically processed, enriched, and restructured to align with the nuanced requirements of applications or consumer use cases [5]. An instance of this can be, for example, personalized investment advice, where the data is gathered from different sources, cleaned, and enriched so that it can be enhanced with predictive models to get precise and actionable insights [5].

Defining and understanding data domains and archetypes are crucial for efficient and well-governed data management. Failing to establish the data domains clearly can lead to a cascade of organizational challenges that undermine data quality, reliability, and accessibility. Organizations risk entangling their data landscape without a well-structured approach to defining and managing data domains and archetypes. This is especially

important for financial organizations, where the number and complexity of data domains can be exceptionally high due to the intricate nature of financial products and services, and it is imperative to ensure compliance with strict regulatory requirements. Therefore, accurately defining and understanding data domains becomes even more critical.

2.3 Financial Service Providers

The financial sector forms a part of the economy consisting of companies and institutions offering financial goods and services to businesses and individual consumers [10], [24]. The financial sector comprises various types of financial institutions, which can be broadly segmented into two main groups: banks and non-banking financial institutions [10]. This classification reflects the regulatory differences between the two as banks, being more traditional financial intermediaries, are licensed institutions [25] and generally more heavily regulated than non-banking financial institutions due to their critical role in maintaining market stability and consumer protection [26].

Albeit often used synonymously with the financial sector, the financial services sector is distinguished by its focus on providing financial services [10]. This distinction becomes evident when considering the nature of financial goods versus financial services. Whilst financial goods refer to tangible or intangible financial products (such as stocks, bonds, or insurance policies), financial services refer to the actions and activities related to managing these products (such as investment advice, loan processing, or asset management) [10]. As banks offer both financial services and financial goods, the financial services sector is not the same as the financial sector but is rather closely intertwined [10]. For the purposes of this thesis, however, the terms financial sector and financial services sector are treated as the same.

Within the financial services sector, the banking industry consists of financial institutions, or banks, licensed to accept deposits, extend loans, and provide other financial services such as currency exchange or individual retirement accounts [25]. Retail, corporate, investment, and central banks are different types of banks [25]. Depending on their operations, banks can also be divided into traditional or challenger banks. Traditional banks are conventional banks with physical branches that are typically combined with some online services and provide a wide range of financial products and services. In contrast, within the scope of this thesis, a challenger bank is defined as a small or medium-sized institution with a limited or no physical presence that focuses on specific product offerings such as loans.

On the other hand, the non-banking financial industry encompasses a variety of non-banking institutions (NBFIs) that do not hold banking licenses but offer other financial services [12]. These financial services include investment services, money transmission, and financial consulting [27]. Examples of NBFIs include insurance and investment companies, hedge and pension funds, and currency exchanges [27]. These entities operate under a different set of regulatory standards than those in the banking industry [12]. Digital NBFIs can be classified as FinTech (short for financial technology) companies, meaning that they use technology to offer financial services and products [11]. While “FinTech” refers to the innovative use of technology in financial services and “FinTech company” denotes the

enterprises driving these advancements, for this thesis, the two terms are used interchangeably, denoting FinTech companies [11].

Leveraging emerging technological trends in finance, such as artificial intelligence (AI) and big data analytics, can significantly enhance a financial company's competitive edge [28]. The success of these technologies in driving innovation is heavily dependent on the quality and accessibility of data. Joshi et al. [18] also highlight that good-quality and accessible data is imperative for facilitating customers and managing both compliance and risk. By improving the quality and accessibility of data, a data mesh could enable more robust AI and data analytics, leading to sharper, more informed decision-making. This could be crucial for financial service providers looking to stay ahead in a rapidly evolving sector.

3 Related Work

The concept of data mesh has gained considerable traction since Dehghani introduced the term in 2019 [29]. Due to the novelty of the data mesh concept, however, limited peer-reviewed publications are available. The existing body of literature on this topic covers the foundational concepts and practical applications of data mesh and case studies from organizations that have adopted this approach.

Dehghani's book [5], several peer-reviewed [17], [30], [31], [32] and in review [16] publications, and other academic works [33], [34] lay the foundational understanding of data mesh. A conference paper by Machado et al. [17] details the concepts and principles of data mesh and presents it as a solution to the limitations of the preceding data architectures. The paper also examines real-life data mesh implementations by Zalando [35] and Netflix [36] to achieve data sharing among these organizations. For instance, in the case of Zalando, the shift from a data lake to a data mesh included keeping the central data lake storage and adding a metadata layer and data governance [17]. Their centralized processing platform simplifies data processing, allowing different business domains to integrate with the infrastructure easily without needing to understand the technical complexities or involve the infrastructure team in specific use cases [17]. However, as these publications primarily seek to elucidate or demonstrate the concept, in contrast to this thesis, they do not focus on the financial sector nor explore how to define the data domains.

Technical contributions have been addressed in [19], [37], [38] and [39]. In one paper [39], Machado et al. define a high-level domain model and a conceptual architecture for data mesh. The authors offer a clear blueprint of the components of the domain model, such as the mesh catalog and mesh nodes, and delineate specific roles for managing and building the data mesh architecture. Furthermore, the conceptual architecture the authors propose serves to complement this by illustrating how domain components integrate with other data mesh aspects like the self-serve data platform, security mechanisms, and overall infrastructure. The outcomes of [39] are further used by the authors in [19] as the basis for proposing a technological architecture by translating high-level knowledge into a concrete artifact. For this, the authors outline a range of technologies, including Apache Ranger for data security or Google Cloud as the cloud service platform, that are suitable for implementing data mesh components. Complementary to this, Butte and Butte [37] describe the data mesh domain architecture, tracing the data from its ingestion and pipeline processing to its final consumption by the data consumers. The authors also detail how data mesh should be implemented in the cloud by using Amazon Web Services (AWS) as an example. While these papers cover the structural and technological aspects of data mesh and its application, they lack insights from industry specialists, do not provide insights into how to define the data domains, nor focus on the financial sector.

Recent case studies [18], [20] offer detailed perspectives on the implementation of data mesh in two organizations: Saxo Bank and the Norwegian Labor and Welfare Administration (NAV). In Saxo Bank, the study focuses on how data mesh principles were applied to enhance data governance in their infrastructure [18]. This included decentralizing data ownership into individual domains, for example, payments and trading, creating data

quality rules and automated pipelines for quality checks, as well as implementing a data catalog including information regarding the data such as the metadata, ownership, and lineage [18]. Vestues et al. [20] conducted 18 semi-structured expert interviews for the NAV case study to explore the organization's transition from centralized to decentralized data management. A recurring theme in the study was the importance of clearly defined and managed domains, as experts expressed concerns about maintaining various domains and underscored the need for well-documented data from different domains for effective inter-domain collaboration. Although these studies offer valuable insights, with the Saxo Bank case study uniquely discussing data mesh in a financial services context, they are tailored to individual case scenarios. In contrast, this thesis adopts a broader perspective as these studies did not incorporate insights from different organizations [18], [20], consider insights from industry specialists [18], or focus on the financial sector [20].

The gap between theory and real-world application is also bridged by [40], [41], [42], [43], which demonstrate the implementation of data mesh concepts across diverse industry-specific contexts, such as energy, dairy, military, as well as oil and gas industries. In a journal article, Pakrashi et al. [40] explore the use of a data mesh architecture within the dairy industry to unify data across various stages of production and distribution, significantly enhancing data accessibility and aiding decision-making processes. Similarly, a conference paper by Li et al. [41] demonstrates how the application of data mesh in park-level power systems can facilitate improved data interaction and governance, a pivotal element for integrating renewable energy sources and achieving operational efficiency in the energy sector. While these works show the practical value of data mesh in various industries, contrastingly to this thesis, they do not focus on the financial industry.

Other papers focus on specific aspects related to the data mesh, such as the privacy challenges in data mesh due to its distributed nature [44], the research of blockchain-powered metadata catalogs for the data mesh architecture [45], [46], contributions to the understanding of how to design self-serve data platforms in data meshes [47], and a systematic gray literature overview of data mesh [48]. In addition, multiple organizations such as Thoughtworks [49] in [50], INNOQ [51] in [6] and [52], PwC [53] in [54] and [55], LTIMindtree [56] in [57], and Credera [58] in [59] have shared their knowledge or experience regarding data mesh, contributing to a deeper understanding and potentially broader application of data mesh. For instance, INNOQ [51] in [6] guides on implementing the data mesh architecture as well as presents real-world learnings and example specifications for data contracts and data products. However, these works do not explore how to define the data mesh domains [6], [50], [52], [54], [55], [57], [59] or focus on the financial sector [6], [54], [57].

There is an increasing interest in the concept of data mesh and how it can be tailored to meet the needs of a specific industry. While these studies provide an important baseline, they do not provide insights into how to define the domains within a data mesh architecture in any sector, including the financial services sector. This results in a gap in the existing literature and presents an opportunity for research. This thesis aims to fill this gap by conducting interviews with industry specialists.

4 Methodology

This section outlines the research questions proposed and the qualitative methodology, specifically interview preparation, data collection, and analysis, adopted in this thesis with relevant justifications. Table 1 presents an overview of the methodology employed in this thesis.

Table 1. Methodology.

Section	Step	Sub-steps	Input	Output
2.1	Research questions			Research questions
2.2	Interview preparation	1. Interview guide creation. 2. Interviewees selection	Research questions	Interview guide
2.3	Data collection	1. Conducting interviews	Interview guide	Interview recordings
2.4	Data analysis	1. Interview transcriptions. 2. Performing thematic analysis	Interview recordings	Themes of findings

An interview-based approach was chosen as it allows for an in-depth understanding of the subject matter [60], such as the business perspective and use cases. Interviews followed the Empirical Standard guidelines for conducting research in software engineering using qualitative surveys to ensure methodological rigor and the credibility of the research findings [61]. Semi-structured interviews, which are supported by the guidelines, were carried out with industry specialists who, with their direct experience and expertise, can provide valuable real-world context and understanding. This approach aims to ensure that the research is grounded in practical reality and offers relevant, applicable solutions and insights. Therefore, to find the answers to the research questions outlined in Section 4.1, an interview-based approach was chosen. The interviews with industry specialists were conducted in February and March of 2024. Afterward, these interviews were transcribed and cleaned, preparing the data for analysis. Thematic analysis was then performed to identify and examine themes within the data to understand the subject matter comprehensively.

4.1 Research Questions

The thesis is guided by the research questions designed to explore the current practices and the implementation of data mesh in the financial sector, ultimately allowing to explore a data mesh domain structure for financial service providers and explore the value of adopting a data mesh architecture. The research questions are the following:

RQ1. How is data currently managed by financial service providers?

The first research question explores how organizations in the financial sector currently manage their data. Understanding data management within the financial sector helps to set the stage for further analysis of its efficacy and potential areas for improvement.

RQ2. How do current data management practices limit financial service providers in achieving their business needs and objectives?

The second research question aims to explore the impediments of current data management practices for financial service providers. It focuses on identifying the business needs and objectives that drive data management and understanding how the data management practices may be limiting the ability to meet these needs and objectives.

RQ3. How can data domains be defined for financial service providers to facilitate achieving business needs and objectives?

The third research question aims to advance the understanding of how the data domains can be defined to support and enhance the achievement of business needs and objectives. Thus, exploring how the data organization into domains should be done – whether the data domains should be organized by process, product, service, or other criteria.

RQ4. How valuable could the adoption of a data mesh architecture be for financial service providers?

Finally, the fourth research question assesses the value of adopting a data mesh architecture. It explores the potential benefits and challenges, if and how it could help to overcome the limitations of current data management approaches, as well as explores its value for different types of financial service providers.

4.2 Data Collection

This subsection outlines the interview preparation steps, including creating the interview guide and finding interviewees, as well as describes the selected interviewees and how the interviews were conducted.

4.2.1 Interview Guide

A semi-structured interview is a qualitative research method that merges the structured approach of pre-determined open-ended questions with unstructured conversations [62]. In this interview format, the interviewer uses an interview guide with key questions whilst still having the possibility to explore the participants' responses or any new themes or questions further [63], [64]. Such interviews are typically conducted only once with an individual or a group and usually last between 30 minutes to over an hour [64].

An interview guide helps to explore the perspectives of different respondents [64]. Moreover, it maintains the focus of the interview with the intent of finding answers to the research questions [63]. Hence, the interview guide (see Appendix I) was developed based on the research questions defined in Section 4.1 and split into three distinct sections: introductory, research questions related, and concluding questions.

Introductory questions served as conversation starters and provided insights into the interviewee's professional experience and their connection with data. The research questions-related section was further refined into four subsections: RQ1-related, RQ2-related, RQ3-related, and RQ4-related. Each subsection focused on a specific research question, featuring a unique set of questions designed to explore and analyze each topic

thoroughly. For instance, the question "*What limitations or challenges have you encountered with the current data management practices in the financial sector?*" aimed to find an answer to RQ2 to uncover the issues posed by how the data is currently managed. Other examples of questions such as "*What would be the main challenges in moving towards adopting a data mesh architecture for a financial service provider?*" and "*In your opinion, how valuable could the adoption of data mesh be for a financial service provider?*" for RQ4 sought to understand the value, including the benefits and challenges, of adopting a data mesh architecture for financial service providers. Finally, concluding questions were asked to allow interviewees to share any additional relevant thoughts or ideas not addressed in the earlier questions.

4.2.2 Finding Interview Participants

The process of finding interview participants consisted of two primary steps: finding a suitable set of participants for the interview and refining that set based on the selection criteria. These steps aimed to ensure that a diverse range of professionals from the financial sector, with their knowledge relevant to the subject matter, would be targeted.

Identifying an initial suitable set of participants was done using the search of the networking platform LinkedIn [65] and reaching out to companies directly. For the LinkedIn search, different keywords such as "data," "head of data," "data engineer," and "data architect" were used to find the candidates with the most relevant background. Different banks and FinTech companies were contacted to seek out participants with the knowledge and experience to participate in the interview. The focus was on financial service providers with a strong presence in the Baltic region for ease of accessibility as well as to make the organizations more comparable.

In order to ensure a varied and representative sample of participants for the study, a selection criterion was established. This selection criterion, as shown in Table 2, considered the tenure and experience of the professionals and aimed to ensure a diverse spectrum of representation of different roles and companies across the sample.

Table 2. Interviewees selection criterion.

Criteria	Description
Tenure	Professionals with longer experience in data architectures were preferred as they were more likely to be able to provide insights into historical and emerging trends.
Experience	Professionals who had hands-on experience designing or implementing data architectures were preferred. A candidate was considered extremely well-suited when they had worked at different financial organizations.
Diverse representation	Attention was given to the diversity of candidates to ensure a representative sample of participants. A preference was given to specialists from different financial service providers to ensure a wide range of expertise and perspectives.

46 selected candidates were approached via email or LinkedIn, depending on which was most suitable, to initiate a connection and extend an invitation to participate in the interview.

4.2.3 Interviews

The nature of the research questions calls for an approach that can adapt to the expertise of the interviewees and the complexity of their insights. Harrell and Bradley [62] highlight that semi-structured interviews are particularly effective for delving deeply into a subject as they allow for a comprehensive understanding of provided responses and leveraging the specialized knowledge of an individual. Hence, when determining the methodology for data collection in this study, the semi-structured interview approach was chosen as it allows for diving deeper into the topics of data management and architecture. This is especially important since it requires specific knowledge and the ability to contextualize that knowledge within the broader industry landscape.

In total, ten interviews with the selected professionals were conducted. The interviews were conducted primarily in English to maintain consistency. However, to accommodate the preferences of one participant, the interview was conducted in Estonian as requested and subsequently translated into English to align with the rest of the data collected. Table 3 describes the interview participants. Specifically, it elicits their background, role, and how many years of experience they have both overall and in their current role. Participants were selected from a range of different types of financial organizations, different organizations, and different roles to include a wider range of perspectives. For example, five participants (I-01, I-02, I-03, I-05, I-09) represented the traditional banking organizations. Moreover, three participants represented challenger banks (I-04, I-06). Lastly, FinTech companies were represented by three participants (I-07, I-08, I-10). Many participants, such as I-06, I-08, and I-10, also had previous experience with other financial companies. All participants were from different companies, with the exception of I-01 and I-05, who, with their different roles and varied backgrounds, ensured a unique perspective nonetheless. Participants' names and other identifying information, such as their organization, were excluded to maintain confidentiality and protect their privacy. This also helped to ensure that the insights they shared could be discussed openly without concern for professional exposure.

Table 3. Participants of the interviews.

Code	Role	Background	Financial institution type	Experience (years)	Experience in role (years)
I-01	Head of Digital Sales and Automation	Responsible for Baltic digital sales and sales automation. Experienced in working with analytics, strategy, customer service, and various data-centric operations.	Traditional bank	14	3
I-02	Head of Data and Business Intelligence	Responsible for both IT and non-IT areas of data, everything related to data governance, data management, and the IT aspects of data.	Traditional bank	15	4
I-03	Head of Data Quality	Leads Data Quality Management and is responsible for defining the policies, frameworks, and guidelines for data management.	Traditional bank	11	6

Code	Role	Background	Financial institution type	Experience (years)	Experience in role (years)
I-04	Head of Data Warehouse Unit	Experience in building and maintaining data warehouse systems, with a focus on coordinating data systems in the financial sector.	Challenger bank	9	3
I-05	IT Chief Architect	Drives technology and architecture in the Baltics and is responsible for a variety of IT-related domains.	Traditional bank	29	4
I-06	Head of Analytics	Leads the data unit, responsible for a broad range of tasks from database management to different analyses.	Challenger bank	15	1
I-07	Head of Data	Leads data warehousing, analytics, governance, and architecture.	FinTech	29	3
I-08	Senior Data Infrastructure Lead	Leads technical direction and vision, overseeing data technologies and security with multiple teams.	FinTech	12	3
I-09	Lead Software Engineer	Responsible for the entire software development lifecycle within the wealth management domain as well as has experience with managing data and building data pipelines.	Traditional bank	19	5
I-10	Staff Software Engineer	Oversees cross-team architecture, works closely with UI and financial teams, managing data for business-to-business products.	FinTech	25	4

The interviews were held online via Zoom [66] individually with each of the participants. Zoom was chosen due to its ability to record meetings and ease of access, enabling participants to join from various locations. Recordings were saved to the cloud, and after the interview, the audio was downloaded to the author's computer locally. Interviews had no direct benefit for the participants, and prior to conducting interviews, consent to record, transcribe, and analyze the interview was obtained from all participants, ensuring ethical compliance and data protection.

The interviews followed a consistent format, beginning with the researcher's introduction, followed by an explanation of the research objectives and an outline of the interview structure. Subsequently, the interview continued by following the prepared questions from the interview guide with the flexibility to ask other questions that emerged from the answers of the participants. During a couple of instances, some interviewees could only share knowledge at a high level due to the restrictions of their employers regarding information sharing. The interviews concluded by asking the interviewees for any additional points they wished to highlight and thanking them for their valuable contribution and participation in the study.

Following the first interviews, some questions in the initial interview guide were slightly modified to make the questions more understandable and precise for the participants. For instance, the question "*What have been the primary business objectives guiding data*

management in the financial sector?" was revised to "*What have been the primary business needs and objectives guiding data management in the financial sector?*" to consider the needs of a business, such as compliance to regulatory requirements, that are alongside but separate from the business objectives. Another question on data domains was rephrased to explicitly mention the word "domains." The question "*What would be an effective way to organize data to support business operations?*" was changed to "*What would be an effective way to organize data into domains to support business needs and objectives?*". This was prompted by the participants' confusion over the term "organization," which they did not associate with domain structures. Additionally, the term "business objectives" was again broadened to "business needs and objectives" to capture a wider range of factors influencing data organization.

The first eight interviews were conducted in February 2024, while the final two took place at the beginning of March 2024. A detailed schedule of these interviews can be found in Appendix II. The duration of the interviews was between 43 and 88 minutes. On average, an interview lasted around 64 minutes. This is consistent with the guidelines suggested by [61], which advise that interviews should range from 30 minutes to over an hour to ensure thorough discussions.

4.3 Data Analysis

The collected data was analyzed using thematic analysis, which Braun and Clarke [13] define as a qualitative data analysis method to systematically identify, analyze, and report themes within the collected data. Thematic analysis was chosen for its flexibility, appropriateness for exploring the novel concept of data mesh without the constraints of a predefined specific theoretical framework, and ability to understand in-depth the impact and nuances of data management in different organizations [13].

As this research employed the inductive or data-driven approach of thematic analysis, the themes emerged from codes that were derived from the data itself [13]. In essence, codes are labels to tag specific segments of data that appear significant or relevant to the research question [13]. From the grouping of related codes emerge themes, which articulate the main ideas emerging from the data around which the data can be discussed and interpreted [13]. The process of thematic analysis was done iteratively as the analysis took place after each interview. Due to this, the codes and themes also developed iteratively, with some codes evolving into main themes, some into sub-themes, and some discarded if they did not fit the evolving thematic structure.

To conduct a thematic analysis, it is essential to have the data in textual format [13]. As there are no strictly established guidelines to adhere to when producing a transcript for thematic analysis [13], the recordings were automatically transcribed using a specialized transcription software called Otter.ai [67]. This software was chosen for its ability to allow users to listen to and see the same parts of the transcription simultaneously and make corrections, which made the transcription process convenient. For one interview, however, the transcription was done manually as it had to be first transcribed from Estonian and,

afterward, translated to English. After the initial transcription of the interviews, the six-phase thematic analysis process [13] was followed.

The first step of the process was *familiarization with the data*. This consisted of repeated reading of the transcripts. Firstly, the transcripts were read and compared to the audio recordings to ensure that the transcript would not have any grammar mistakes and conveyed the ideas true to their original nature. For example, punctuation had to be correct so as not to alter the meaning of the data. Then, active reading of the data was done to understand the data thoroughly and identify ideas for codes and themes.

Secondly, an *initial set of codes* was generated. The codes "identify a feature of the data that appears interesting to the analyst" [13:18]. This was done using the MAXQDA tool [68], which can be used for qualitative data analysis and provides specialized tools for inductive coding approaches [68]. It is a convenient solution for creating, visualizing, and hierarchically organizing a code system [68]. As it also supports direct imports from Otter.ai, the transcriptions were conveniently imported into the software. In the initial coding phase, each interview was reviewed to assign the relevant codes based on the RQs. Following Braun and Clarke's [13] suggestions, the coding strategy aimed to capture as many potential themes as possible and include expansive data extracts, acknowledging that some data might be coded multiple times if pertinent. The coding was refined with each interview – new codes were introduced, and some were combined or categorized as subcodes, leading to 71 unique codes. The five most frequent codes are detailed in Table 4.

Table 4. Five most frequent codes from thematic analysis.

#	Code	Frequency	Interviewees mentioned
T1.1.3	Data consolidation to data warehouse	10	I-01, I-02, I-03, I-04, I-05, I-06, I-07, I-08, I-09, I-10
T2.1.1	Compliance	9	I-01, I-02, I-03, I-04, I-05, I-07, I-08, I-09, I-10
T8.3	Value depends on the organization	8	I-02, I-04, I-05, I-06, I-07, I-08, I-09, I-10
T1.1.5	ETL pipelines	7	I-03, I-05, I-06, I-07, I-08, I-09, I-10
T1.1.2	Ownership of data	7	I-01, I-02, I-04, I-05, I-06, I-09, I-10

Consequently, the third step consisted of *sorting the generated codes into themes*. This meant that codes that shared similarities were clustered to determine if they could logically converge into a theme. This was done using the affinity diagram method as it allows for enhanced visualization and effective arrangement and re-arrangement of the codes and interview excerpts using sticky notes. For instance, codes "Unclear data ownership" along with codes "Lack of data competence" and "Chain of command" were grouped into the sub-theme "Culture" to capture the challenges related to organizational practices, attitudes, and competencies.

The fourth step was the *refinement of themes*, which encompassed refining the candidate themes that were developed. A rule of thumb is for themes to be coherent and clearly

distinguishable from one another [13], as well as support finding answers to the research questions outlined in Section 4.1. During this phase, it became clear which themes lacked sufficient supporting data or were too varied, requiring either a merger with other themes or a division into new themes.

Finally, steps five and six *named the themes* and *produced the final report*, respectively. Upon refining the themes, each theme was assigned a name to capture the core concept of the theme as well as a description to provide an overview of its content and scope. Then, a final analysis and report was produced. In total, nine themes and fourteen sub-themes were created (Figure 3).

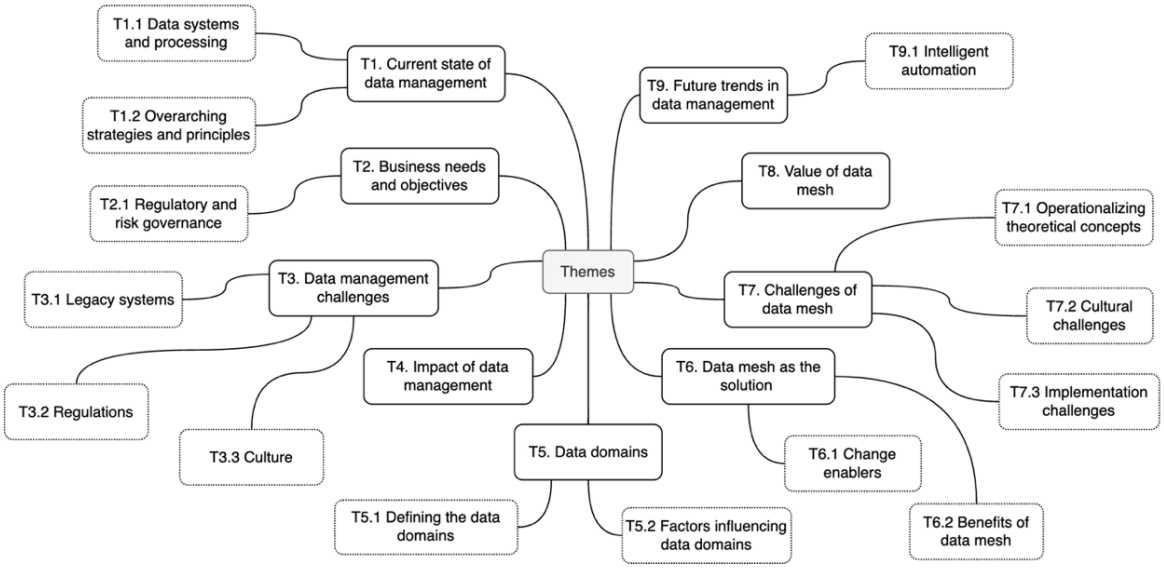

Figure 3. Themes from the interviews.

The full list containing the themes, subthemes, codes, and examples can be viewed in Google Sheets [69].

5 Results

The following section aims to give an overview of the results and how they relate to the research questions outlined in Section 4.1. This section is structured into four subsections, each corresponding to a specific research question. The first subsection outlines how the data is currently managed by financial service providers (RQ1). Following it, the second subsection will focus on the business needs and objectives and how the current data management practices limit the ability of financial service providers to achieve those (RQ2). In the third subsection, results are given on how the data domains could be defined for financial service providers (RQ3). Lastly, subsection four presents the findings regarding the value of data mesh for financial service providers (RQ4).

5.1 Data Management

The interviews determined how data is managed by financial service providers (RQ1). This varies between traditional banks, challenger banks, and FinTechs, touching on the following components depending on the type of the organization: monolithic source systems, microservices, offline and online delivery of data, data warehouses, data lakes, data criticality, data ownership, and data governance specialists. The results are presented based on the type of financial service provider in the following order: traditional banks, challenger banks, and FinTechs.

The lifecycle of data for traditional banks starts with the "*monolithic [source] systems*" (I-05), as mentioned by interviewees I-01, I-02, I-03, I-05, and I-09. These systems were also referred to as "core banking systems" (I-02) and "big monolithic systems" (I-05). This means that when a customer registers at a bank, that information is recorded in the transactional source systems and the non-transactional systems, such as the "CRM systems, where we are storing customer data" (I-02). Once recorded in these initial systems, data then flows into operational databases (I-03, I-05, I-09). For traditional banks, the "many operational databases" (I-03) can often be part of the larger monolithic system that supports a wide range of banking functions. Interestingly, I-05 also mentioned using a database platform: "If you are talking about relational database as a platform, our core banking system is built from OpenACC database platform."

Afterward, data is transferred and stored in analytical systems. *Offline data movement* often relies on ETL pipelines (I-03, I-05, and I-09). Interviewee I-03 noted that "For the majority of cases, it is still point-to-point ETLs." Interviewee I-05 echoed this, highlighting the batch processing nature of this mechanism: "ETL process is used to load data on a daily basis." Moreover, interviewees (I-05 and I-09) also mentioned the use of *online or near real-time data delivery systems* to other databases or analytical systems. As mentioned by I-05, "We send data using Kafka […] This is done for analytical purposes, but we have demand to get it quicker." Furthermore, I-09 also mentioned that "You can use Kafka as a persistent thing, but we're using message brokers that are only meant for data movement." The replication of data can, therefore, happen in near real-time or in batches, depending on the system's requirements. Data is typically further consolidated into the *data warehouse* (I-01, I-02, I-03, I-05, I-09). For instance, I-09 acknowledged that there is a "centralized data warehouse

solution". Notably, interviewees I-02, I-03, and I-09 also mentioned the existence of "many warehouses" within their institution. The interviewees described two use cases for having these data warehouses. As emphasized by I-02, data warehouse solutions are "needed for regulatory and for management reporting." I-03 expanded it, explaining to use it for "regulatory compliance and to cater the risk needs or risk compliance. […] getting the insights for their own business performance and business decision making" and pointed out a benefit of using this data architecture by noting, "When you're getting the data from the single consolidated view, […] it makes the reports consistent." Another analytical storage repository, discussed by the interviewees (I-02, I-03), is the *data lake*. I-02 described it as a more "modern data infrastructure" that serves as "an intermediate component to have the same data structures as in the source systems. […] Where you don't need data to be consolidated and well-structured". I-02 further elaborated that data lakes are often used for "different kinds of data science use cases," including the development of predictive analytics models such as "propensity model or churn model" to understand customer behaviors and to forecast trends.

Notably, when describing how data is managed, interviewees I-02, I-05, and I-09 emphasized the *critical* nature of *data* as being an asset that requires control and governance. For instance, I-02 asserted that "data is a critical asset that should be managed in a well-controlled manner. This is far away from being nice to have." I-09, specifically, highlighted the minimum principle of data: "We don't store more information than needed because you can store only as much data that you need." Furthermore, the interviewees (I-01, I-02, I-09) also touched upon the topic of *data ownership*. Significantly, I-02 stated that domains themselves own and govern the analytical data: "It means that in every business line, you have the data owner - data owners in retail, in corporate, in finance, in risk, etc. We don't have one centralized team of data owners who are responsible for everything. […] my preference is always so that the business process owner is also the data owner." This contrasts interviewees I-01, I-05, and I-09, who recognized the existence of centralized data ownership. For instance, I-09 highlighted that "the data warehouse is centralized and on-premises. […] The mismatches are checked in the background by the data warehouse team." Interviewee I-02 also highlighted having *data governance specialists* such as "a data governance officer" and "a team of data stewards, who are responsible that data quality is improving."

Challenger banks were represented by interviewees I-04 and I-06, who also underscored that the lifecycle of data starts in the source systems, where customer interactions are initially recorded and processed as those are the "systems with which end users are operating" (I-06). Interviewee I-04 described it, referring to a *monolithic* architecture: "Customer goes to the [bank's] page and makes a request there. Applications have their own module that goes to the request database." Both interviewees also mentioned operational databases, with I-04 specifying, "There are many databases, and they are of different types." Interviewee I-06 also touched upon *offline data delivery* using "data pipelines or the ETL processes." Both I-04 and I-06 noted *data warehouses* for consolidating data, with I-06 highlighting the use of multiple warehouses. The interviewees, however, did not highlight any unique perspectives compared to the interviewees from traditional banks. Furthermore,

interviewee I-06 also touched upon *data lakes*, noting that "The data lake concept, we are using very seldom." Regarding *data criticality*, specifically adhering to the minimum principle when it comes to data collection, I-04 introduced regulations as the reason for it, explaining that "In the financial sector, the GDPR is so strict that you can't pick up data arbitrarily […]. Rather, the minimum principle is to collect only the data you know you need." Additionally, both interviewees mentioned *centralized data ownership*, with I-06 noting, "We are people of six to support everything."

Contrastingly, FinTech companies break away from the monolithic source systems in favor of "using *microservices*" (I-07), as denoted by representative interviewees I-07, I-08, and I-10. This microservice architecture breaks down the monolithic model into smaller, interconnected services, each focusing on a specific business capability [70]. Interviewee I-07 reasoned that microservices can foster a more resilient system architecture, "if the application's functionality is inside of one system only, then it makes developing this huge monolith very risky." Each microservice may interact with its own operational database. For instance, I-08 described it as "service has the service database, which has like an internal representation of those interactions to the customer."

All FinTech interviewees mentioned ETL pipelines as a means for *offline data delivery*, with I-07 specifically contrasting it to online delivery mechanisms: "If it's ETL, then it's not so real-time." Similarly to the traditional bank interviewees, both I-07 and I-08 also noted Kafka as an *online data delivery mechanism*. For instance, I-07 mentioned, "We are using Kafka as well. Kafka processing is closer to real-time". The FinTech interviewees also mentioned *data warehouses* and the two reporting use cases mentioned by both traditional and challenger banks. Notably, I-10 introduced other use cases: "very frequent use is data discovery or some ad hoc investigation." The interviewee (I-10) also touched on using *data lakes* from a different perspective, implying that while data lakes are valuable for their ability to centralize "raw, unformatted data," they can become a "problem from the contract perspective."

Both I-07 and I-10 also highlighted the aspect of *data criticality*, with I-07 (likewise to I-04 from a challenger bank) asserting that data must be "protected and managed according to our regulations." Furthermore, I-10 discussed *data ownership*, highlighting the transition toward decentralized operational data management, yet noted the centralization of analytical data. The interviewee remarked, "It's microservices, where each team is managing their own [operational] data. […] Analytical data centralized into a data warehouse. We use Snowflake".

Table 5 provides an overview of the data management components as discussed by interviewees, reflecting the distinctive perspectives of traditional banks, challenger banks, and FinTech companies.

Table 5. Current data management.

#	Financial institution type	Data management component	Result summary	# (Interviewees mentioned)
1	Traditional bank	Monolithic source systems	Expansive legacy core banking systems that store and process transactional data.	5
		Offline delivery of data	Batch replication of data via ETL pipelines or message brokers from source systems to other systems, such as data warehouses.	3
		Online delivery of data	Near real-time replication of data, using technologies such as Kafka, from source systems to other systems.	2
		Data warehouses	Central data repositories that are used for consolidating data for regulatory and management reporting.	5
		Data lakes	Central data repositories that are used less frequently than data warehouses. Mainly to hold unstructured data and for data science use cases.	2
		Data criticality	Recognizing data as a critical asset that must be managed in a controlled manner.	3
		Decentralized data ownership	Data domains are responsible for their analytical data.	1
		Centralized data ownership	Responsibility for analytical data falls upon the centralized data team.	3
		Data governance specialists	Centralized data specialists responsible for data governance such as data stewards.	1
2	Challenger bank	Monolithic source systems	Legacy systems where customer interactions are initially recorded and processed.	2
		Offline delivery of data	Data is replicated in batches via ETL pipelines to external systems.	1
		Data warehouses	Centralized analytical data management is overseen by a central data team.	2
		Data lakes	Used rarely for specific unstructured data storage needs.	1
		Data criticality	Emphasizes minimal data collection in compliance with data protection laws.	1
		Centralized data ownership	Centralized analytical data management and a central data team is responsible.	2
3	Fintech	Microservices	Monolithic system functionalities are broken down into microservices to enhance resilience and scalability.	3
		Offline delivery of data	ETL pipelines for replicating batch-oriented data to external systems.	3
		Online delivery of data	Near real-time replication of data to external systems. For example, data streaming using Kafka.	2
		Data warehouses	Central for data analytics and reporting, supports ad hoc investigations.	3
		Data lakes	Central repository used for storing raw, unformatted data.	1
		Data criticality	Data is protected and managed according to the regulatory frameworks.	2
		Centralized data ownership	Analytical data is centralized, contrasting with decentralized operational data management.	1

5.2 Impact of Data Management

In this section, results for how data management practices limit the ability of financial service providers to achieve their business needs and objectives (RQ2) are presented. From the interviews, three key aspects emerged. The first aspect outlined business needs and objectives, which refers to what financial service providers are required to do and aim to achieve. The second aspect focused on the challenges that arise when businesses strive to meet these needs and objectives. The third aspect examined if and how these data management challenges affect the business needs and objectives that data management is supposed to support. Consequently, the results are presented in the same sequence: business needs and objectives, data management challenges, and the impact of data management challenges on business needs and objectives.

5.2.1 Business Needs and Objectives

Business needs are externally imposed critical regulatory requirements that financial service providers must satisfy, whilst business objectives are the internal goals these organizations strive to achieve. In the interviews, compliance and risk management emerged as business needs. The interviews also identified the following business objectives: improving financial performance, enhancing customer experience, having quality data, creating process efficiency, and ensuring scalability to support growth. The needs and objectives are organized according to the category of financial service provider (traditional banks, challenger banks, and FinTechs) and are further ordered by the frequency with which interviewees mentioned them, beginning with the most cited business need or objective.

All traditional bank interviewees (I-01, I-02, I-03, I-05, I-09) mentioned *compliance* as a need for the business. I-05 placed it as the top priority: “Number one is compliance. […] If you’re not compliant, you will lose your banking license.” The interviewee (I-05) explained these extensive compliance demands, comparing “banks to other types of industries, it is an extremely regulated industry. There are quite many regulations coming from local authorities, from the European Central Bank. […] GDPR is one example. DORA, it’s another regulation.” Adding to this further, I-09 also underscored the impact of regulations on data retention, denoting that there are “specific requirements on how long you must store those and how you would store it. It’s very specific, and in some cases, 10 years plus you must maintain those.”

The interviewees (I-01, I-02, I-03, and I-05) also noted that *enhancing customer experience* is one of the objectives of traditional banks. For instance, I-02 revealed that “it’s all about making sure that we are there to support our customers. To create additional value for the customers”; however, equated it to being natural when it comes to “business as such.” Interviewees I-01, I-03, and I-05 additionally touched upon personalized offers as a way of enhancing customer experience. Respectively, both I-01 and I-03 shared that they are capturing data and are calculating “in the background automatically all the time, what’s the best fitting service plan for you” (I-01) and “what is the next best action for that customer” (I-03).

Akin to compliance from the data governance perspective, three traditional bank interviewees (I-02, I-03, I-05) also highlighted the need to *manage risks*. For instance, I-02 emphasized the importance of maintaining trust and managing risks effectively, stating, "We need to make sure that we are trustworthy. [...] This BCBS 239 requires that we manage our data well, manage the risks well."

Moreover, *quality data* was another aspect interviewees I-02, I-03, and I-09 underscored. "Then, of course, what is important is the quality of data." (I-09). I-02 equated it to the "most important aspect" and highlighted its importance for decision-making, stating, "Using bad data, you could make bad decisions." Three interviewees (I-01, I-03, I-09) also identified *creating process efficiency* as an objective. Interviewee I-03 encapsulated this objective: "We're focusing on how we can cater to more use cases with less effort." Similarly, I-09 commented on the importance of "speed to get a new database system to the data warehouse" and its criticality in supporting the launch of new products. *Improving financial performance* was highlighted as an objective by I-01 and I-02. For instance, I-01 mentioned focusing on "increasing digital income" to enhance the bank's revenue streams through digital channels.

From the perspective of challenger banks, interviewee I-04 also mentioned *compliance* as a business need but did not present a viewpoint that differed from those expressed by interviewees from traditional banks. The interviewee remarked that they are "influenced by the regulations of various central banks and financial inspections, which the bank must comply with." The same interviewee (I-04) also pointed out the need to *manage risks*, stating that they do "risk modeling based on the data in the warehouse."

Moreover, *improving financial performance* was highlighted as an objective by I-04, who noted: "The main goal is still sales and gaining market share." Furthermore, I-04 identified that another objective for the challenger bank is to *enhance customer experience*. The interviewee brought in another distinct perspective, mentioning automated decision-making as a way to enhance customer experience: "The fastest loan decisions are made in less than half a minute. […] Even if they get a negative decision, it still saves them time." The second challenger bank interviewee, I-06, underscored *creating process efficiency* as the sole objective. I-06 articulated that the essence of it is to do "everything now and faster," aiming to empower business users to become more self-reliant through better tooling.

Compliance was also identified as a business need by all FinTech interviewees (I-07, I-08, and I-10). Interviewees I-07 and I-08 focused specifically on regulatory reporting. I-07 outlined, "For the financial sector, one thing that is guiding all are regulatory reporting requirements." Interestingly, whilst I-08 agreed that compliance is a need, underscoring the additional "regulations concerning data retention," the interviewee pointed out that "collecting customer data, private information […] It's not unique to banks. [Businesses from other industries] also do verification and know your customer. There are also regulatory requirements governing how data should be handled […] In that sense, the governance mechanisms between a FinTech and other industries, where you still collect customer data, there isn't much difference."

Furthermore, two interviewees (I-08 and I-10) emphasized *improving financial performance* as an objective. For instance, I-10 uniquely brought attention to the importance of managing costs as otherwise, it is easy to "drive up the cost of your platform" and, by extension, impact the company's financial performance. Interviewee I-07 also pointed out *risk management* as a need, referencing it in terms of "a lot of finance risk-related processes" present in the sector. Additionally, I-10 acknowledged having *quality data*, stating it as a "requirement that your data should be clean, correct, up to date," but referred to it as "almost implicit."

Interestingly, a FinTech interviewee discussed a unique perspective that was not mentioned by traditional or challenger banks. I-08 mentioned *scalability* as a pressing requirement, as FinTechs are often on an upward growth curve. According to this interviewee, this contrasts banks, who commonly focus on retaining their "existing market share" and whose "data is not growing that much year over year." As I-08 explains, "Growth brings continuous pressure on systems," underlining the need for an infrastructure that can adapt and scale.

Table 6 outlines the business needs and objectives of financial service providers that can have an effect on data management.

Table 6. Business needs and objectives.

#	Financial institution type	Business need/objective	Result summary	# (Interviewees mentioned)
1	Traditional bank	Compliance	Complying with regulations, such as GDPR and DORA, is paramount as otherwise, banks might lose their banking license.	5
		Enhancing customer experience	Banks aim to create additional value for customers through personalized services and recommendations.	4
		Risk management	Trust and risk management are critical, guided by standards such as BCBS 239 for data and risk handling.	3
		Quality data	The quality of data is essential for making informed decisions and ensuring accuracy in reporting.	3
		Creating process efficiency	Efficiency in processes is targeted to serve more use cases with less effort, enhancing speed and agility.	3
		Improving financial performance	Enhance financial performance through, for example, digital channels.	2
2	Challenger bank	Compliance	Like traditional banks, challenger banks must adhere to regulations imposed by central banks and financial inspections.	1
		Risk management	Risk modeling is critical for challenger banks and is often based on data from warehouses.	1
		Improving financial performance	The primary goal is sales and increasing market share.	1
		Enhancing customer experience	Customer experience is improved through automated decision-making.	1
		Creating process efficiency	The focus is on achieving more with less time using better tooling, aiming for the empowerment of business users.	1

#	Financial institution type	Business need/objective	Result summary	# (Interviewees mentioned)
3	Fintech	Compliance	Providing reports to regulatory institutions and complying with industry-specific data retention requirements.	3
		Improving financial performance	Increase market share and manage costs.	2
		Risk management	Engaging in many risk-related processes.	1
		Quality data	Clean, correct, and up-to-date data is almost an implicit requirement.	1
		Scalability to support growth	Infrastructures have to be scalable to accommodate rapid growth and data volume expansion.	1

5.2.2 Data Management Challenges

The interviews revealed challenges that financial service providers face related to data management. The following challenges were identified: complying with regulations, cardinality and complexity of systems, maintenance efforts, data governance challenges, data treated as a by-product, data migration challenges, lack of real-time data, and culture-related challenges, that are unclear data ownership, lack of data competence, and chain of command. The challenges are organized according to the type of financial service provider (traditional banks, challenger banks, and FinTechs) and the frequency of mentions by the interviewees for each organization type.

The most prominent challenge for traditional banks revealed from the interviews was *complying with regulations* (I-01, I-02, I-03, I-05, I-09). These compliance efforts were described as significant. For instance, I-05 emphasized, "Some years ago, we calculated that 50% of development capacity in the organization was spent on compliance," siphoning focus and resources away from data management for financial organizations.

Legacy systems typically have *high cardinality and complexity*, being mentioned as a challenge by interviewees I-01, I-02, I-03, I-05, and I-09. Interviewee I-02 provided insight into this: "It sounds quite straightforward and easy, but in reality, you have hundreds of source systems, several data warehouses, several different systems. […] it's much more difficult and complex." I-02 and I-08 (FinTech interviewee considering it from the perspective of a bank) both identified this complexity to be, at least to some degree, a result of non-organic growth. For instance, I-08 stated: "Usually why banks […] have a messy set up is also because they've gone through a lot of acquisitions." This complexity, in turn, *increases the maintenance effort* (I-02, I-03, I-09) of the systems. I-02 described it as requiring "much more time ensuring the basics," whilst I-03 and I-09 focused on the effort of integrating data sources with existing systems. I-03 explained that there is a "lack of reusability of the same data out interfaces […] That is to some degree limiting our progress or our adoption of the newest or more practical data management aspects."

Issues related to *data governance* were noted by interviewees I-02 and I-03. As I-02 pointed out, the challenge encompasses the broader realms of "implemented data governance and data management processes" and identified not having "an aligned data model" as a challenge. Moreover, I-03 remarked that the data owners do not know "who the consumers

of the data are." The interviewees (I-01, I-09) also brought up issues regarding *data being treated as a by-product*, specifically regarding its quality and accessibility. "It's the place where we come from. It's legacy systems that haven't been meant for data collection or accessibility" (I-01).

Furthermore, interviewee I-05 addressed the *challenges of data migration* in the case of legacy systems. The interviewee discussed that for the "big monolithic systems, it's difficult to refactor anything." I-08 (FinTech) also discussed this from the perspective of traditional banks, emphasizing the complexity of maintaining parallel operational paths during the transition to ensure business processes are not disrupted: "You need to continuously support parallel paths or ways of doing things as well." Significantly, I-05 brought to light a specific scenario necessitating migration as "banks, typically, run banking systems on-premises" and must redesign their systems to enable operations to alternate between on-premises and cloud services according to requirements, describing it as "quite an expensive refactoring."

From a data culture perspective, interviewees I-01 and I-03 mentioned *unclear data ownership* as a challenge. "No ownership of data at times, or the owners are there, but they don't know or understand what they own." (I-01). Compounding the issue of ownership is the *lack of data competence,* as highlighted by I-02 and I-05. For instance, interviewee I-05 expressed concern, "If you're talking about data management, from my point of view, the biggest challenge is competence." Cultural challenges were also identified by I-05 and I-08 (FinTech from a traditional bank's perspective), albeit from a wider organizational culture perspective. Both interviewees underscored the *chain of command* being a challenge. I-08 explained it as: "If the command chain stays strictly hierarchical and you need to go five levels up to explain somebody deeply why something's needed, they're so distant from it that there will never be any alignment or agreement. That is why there's a risk of stagnation in some places." Moreover, the challenge of *not having near real-time data* was emphasized by interviewees I-01 and I-09. Specifically, I-01 pointed it out as a significant gap in their data management capabilities: "What we are in most cases missing is real-time data."

The interviews also revealed that, similarly to traditional banks, challenger banks also experience challenges related to legacy systems. *Increased maintenance effort* of the systems was mentioned by both I-04 and I-06. I-06 described it as: "You will reach a point where you're spending almost all your time maintaining these queries and processes instead of delivering new insights." Both of the interviewees also mentioned *data being treated as a by-product*. I-06 described that "these legacy systems bring about less quality data and isolation of data," whilst I-04 commented that field of data is "always in the background."

Interviewee I-04 also highlighted *complying with regulations* as a challenge, noting that if there are new regulations "from the central bank, it may happen that you have to rebuild things." I-04 also pointed out the issues regarding *data governance*. Similarly to I-02 from a traditional bank, I-04 discussed challenges related to not having aligned data models, explaining that "if there are many domains, the content of the data may be the same, but the name may be different" and related data governance to being "quite a chaotic issue."

Moreover, interviewee I-06 outlined the *data migration challenges*. In particular, I-06 explained that they are giving out "long-term contracts (10, 20, 30 years long) such as loans

or insurance contracts" and emphasized that "you can't forget about those old contracts and consider only the new ones. Those are financial obligations that you have. There are very strict rules on how you can migrate or handle those from the legal side. [...] Usually, those migrations might be very complex and tricky". Interviewee I-04 also mentioned a cultural challenge regarding *the lack of data competence*. The interviewee highlighted, "We have a relatively small bank, and even we have a very difficult time getting these product team developers to go down to the data level."

FinTech interviewees did not mention many challenges. Both I-07 and I-10 mentioned challenges related to *data governance*. I-07 identified "ensuring the quality of documentation" of data objects as a challenge since "during the lifecycle, [the data object] has been expanded and maybe the original documentation is not accurate and valid anymore." I-10, echoing a concern similar to that of I-03 from a traditional bank, stressed a lack of awareness about "who the users might be" and "what information may be needed."

I-07 also highlighted *complying with regulations* as a challenge. The interviewee denoted that the "Financial sector is highly regulated. Regulations, particularly finance-related regulations, are different in different countries. One thing that can drive all those processes is how that system has to be developed and managed." Interestingly, interviewee I-08, drawing from previous experiences, discussed challenges exclusively associated with traditional bank and did not mention any specific challenges faced by FinTechs.

Table 7 provides an overview of the data management challenges for traditional banks, challenger banks, and FinTechs.

Table 7. Data management challenges.

#	Financial institution type	Data management challenge	Description	# (Interviewees mentioned)
1	Traditional bank	Complying with regulations	Banks allocate significant resources to compliance, affecting development capacity.	5
		Cardinality and complexity of systems	Complex systems resulting from non-organic growth increase management difficulty.	5
		Increased maintenance efforts	Maintenance complexity and integration challenges with existing systems.	3
		Data governance	Implementing consistent data governance and alignment across the organization is a challenge.	2
		Data treated as a by-product	Legacy systems were not originally designed for data accessibility and quality.	2
		Challenges with data migration	Migrating data in monolithic systems requires expensive refactoring and careful parallel operation maintenance.	2
		Unclear data ownership	There is confusion over data ownership or a lack of understanding among owners.	2
		Lack of data competence	Finding skilled professionals in data management is a challenge.	2
		Chain of command	Rigid hierarchical structures can impede innovation and responsiveness.	2

#	Financial institution type	Data management challenge	Description	# (Interviewees mentioned)
2	Challenger bank	Increased maintenance efforts	Maintenance of legacy systems consumes time that could otherwise be used for delivering new insights.	2
		Data treated as a by-product	Legacy systems often result in isolated and lower-quality data due to data being a secondary concern.	2
		Complying with regulations	New regulations can necessitate changes in system architecture and operations.	1
		Data governance	The absence of unified data models poses challenges, with data governance being viewed as a chaotic issue.	1
		Challenges with data migration	Migrating long-term financial contracts must adhere to strict legal considerations and can be complex.	1
		Lack of data competence	Employees struggle to grasp the intricacies of banking data.	1
3	FinTech	Data governance	Maintaining up-to-date documentation and metadata is a challenge.	2
		Complying with regulations	Adherence to varied regulatory requirements across countries.	1

5.2.3 Impediments of Data Management

As indicated by the interviews, how data is managed can affect a financial organization's ability to meet its business needs and objectives. The interviews identified delayed time to market, inefficient resource allocation, inability to achieve business objectives, central data team as a bottleneck, and loss in competitiveness as impediments. Furthermore, some interviewees did not see the data management challenges not seen as impediments. The results are presented based first on the financial organization type and then organized by the frequency of mentions by the interviewees.

Traditional bank interviewees conveyed varying perspectives on how data management impedes achieving business needs and objectives. One such impediment, as noted by interviewees I-01, I-05, and I-09, is delayed time to market. "Sometimes it means that the time to market is long because you need to stitch the data together." (I-01). However, I-09 discussed that these delays impede maintaining visibility in a competitive market: "The timing can't be long. Otherwise, you lose the visibility." Another impediment, identified by interviewees I-01, I-05, and I-09, is the need for *extra resource allocation*. I-01 described it as follows: "That means that we are working hard with implementing workarounds, things we can do in order to have compliant and good level data." I-05 added to this, mentioning that "you need to have overhead people who will work on it. It's extra work."

Furthermore, interviewee I-09 outlined that issues with data management can potentially lead to a *loss of competitiveness*. The interviewee noted, "Otherwise, you are not faster than your competitors and don't have the competitive edge." Interviewee I-05 also discussed that for a centralized data architecture, the *centralized data team can become a bottleneck*, stating, "You have centralized ownership, and they will become a bottleneck." Notably,

interviewee I-01 described that the challenges of data management can lead to businesses being *unable to achieve business objectives*. The interviewee stated: "Sometimes it means that you can't do the business objective at all." This contrasted the opinion of I-02, who did *not consider data management practices to be impediments* to business. The interviewee remarked, "I will say that we have managed to bypass the limitations. […] In general, customer experience and everything else starts with the quality of advisory. You cannot always replace something automated with just good customer service."

A challenger bank representative, I-06, also emphasizes that how data is currently managed leads to the *centralized data team becoming a bottleneck*. I-06 explained that a small central team struggles to cope with the volume of requests: "We are not able to deliver if we would need to fulfill some ad hoc requests and queries." In contrast, I-04 argued that data management *practices do not impede* achieving business needs and objectives. The interviewee shared: "I haven't heard anyone mention that lack of data has affected profitability. It has some influence, but at the same time, you can always look at things from another angle and make the prevailing situation clear by doing the analysis."

FinTech interviewees shared the perspective that data management *practices are not impediments*. Interviewee I-08, however, outlined *loss of competitiveness* as a possible impediment caused by data management practices, stressing that data problems can "decrease the quality in decision-making, pose risks in making the right strategic decisions for the company," leading to a potential "loss in competitive edge, the eventual loss in market share, revenues, everything."

Table 8 summarizes the impediments arising from data management to achieving business needs and objectives.

Table 8. Impediments of data management.

#	Financial institution type	Impediment	Description	# (Interviewees mentioned)
1	Traditional bank	Delayed time to market	Data management complexity can lead to longer product development cycles and delays in launching new services.	3
		Extra resource allocation	Current data management necessitates additional resources, both in terms of human resources and technological investment.	3
		Loss in competitiveness	The inability to efficiently manage data can result in a loss of market visibility and competitive edge.	1
		Centralized data team as a bottleneck	Centralized ownership of data can lead to bottlenecks and inefficiencies in data processing and usage.	1
		Unable to achieve business goals	Data management practices may impede the realization of strategic business objectives.	1
		Current practices are not seen as impediments	Finding ways to work around the limitations, not viewing the current practices as obstacles.	1
2	Challenger bank	Centralized data team as a bottleneck	A small central data team can struggle to handle the volume of data requests, slowing down operations.	1
		Current practices are not seen as impediments	Challenges of current data management practices are not viewed as affecting business profitability and can be counteracted by other measures.	1

#	Financial institution type	Impediment	Description	# (Interviewees mentioned)
3	FinTech	Current practices are not seen as impediments	Issues with current data management are seen as something that can be mitigated, not impediments to achieving business objectives.	2
		Loss in competitiveness	Data issues can potentially lead to suboptimal decision-making, affecting strategic positioning and market share.	1

5.3 Data Domains

This subsection presents results for RQ3: "How can the data domains be defined for financial service providers to facilitate achieving business needs and objectives?" Two main areas with regard to data domains surfaced from the interviews. The first area discussed how data domains could be defined. The second area outlined the factors that influence defining these domains. Therefore, the results are presented following the same logic, starting with outlining the results on defining the data domains and subsequently, the factors that influence defining these domains.

5.3.1 Defining the Data Domains

The interviews identified that the data domains should, on a high level, be aligned with the business domains and can further be defined using product-centric or process-driven approaches. The results are presented first for traditional banks, then for challenger banks, and lastly, for FinTechs. For each, the results are ordered, starting from the most frequent result for how the data domains can be defined.

In the interviews, traditional bank interviewees I-05 and I-09 emphasized that how the data domains are defined should be in *alignment with the business domains*. For instance, interviewee I-09 stated that "Mostly, it's business domains that lead the way how data is structured and stored" and underscored that "every domain consists of many subdomains," where the individual subdomains can make use of "core logic that's shared." Interviewee I-05 described the subdomains to be defined by products: "You should have the data domains aligned with the business domains by default. […] Product-specific business functionality and data should be located in a business domain." The interviewee also noted the benefit of organizing data in this way as it enables "faster time to market" (I-05). A *product-centric* approach was also outlined by interviewees I-02 and I-03, who described this to involve organizing data based on product categories such as "mortgages" (I-02) and "lending" (I-03). I-02 described it as: "the organization of data, or the data domains, typically, the products are more or less the same. […] That's why it's quite easy to define the domains of the financial industry. […] I would define those as per the products." On the other hand, interviewee I-01 proposed a *process-driven* approach to data organization. This approach was also suggested by I-02 as an alternative to product-centric domains. I-02, however, described it as organizing data domains by customer "value streams." While both process-driven and value-stream data domains prioritize the flow of activities, they differ slightly in focus. Process-driven domains center on specific operations, such as loan giving, while

value-stream domains encompass the entire customer journey [71], such as the initial inquiry about a loan to ongoing loan services. This thesis considers the two to be the same.

From the perspective of challenger banks, I-06 also discussed *aligning data domains with business domains*, giving an example that "there can be some rough split of financial data, sales data, and risk-related data." The interviewee also noted that it is important not to create "an artificial split. Because internet sales and physical sales still happen in the same core system. The point of entry might be different, but they're still generating the same kind of data. For example, loan contracts still start to live their lifecycle in the same way." Furthermore, interviewee I-04 proposed defining *process-driven* data domains, discussing that it should be aligned with the "lifecycle of products", such as the distinct processes of "getting a loan" and "managing a loan."

Interviews with participants from FinTechs outlined an inclination towards a *product-centric* structure (I-07, I-08), with I-07 describing this way of defining data domains as "quite common and widely used." Moreover, interviewee I-08 considered it as natural alignment: "If the product teams own the services that gather the data and have designed the data model […] It just works". The interviewee (I-08) also mentioned that defining the domains should *align with the business domains* and equated the business domains to be further divided by products: "How I think, how we think and how I see people think about splitting the data domains, it's more splitting them by business and product domain, that's the split." Contrastingly, another FinTech interviewee, I-10 focused on *process-driven* data organization but equated the processes to "services," approaching it from the perspective of microservices. The interviewee explained that the data domains should be organized according to "the service level that's being offered. If you have a loan-giving service, then usually, that is one set of data. And that's usually grouped together. […] It's not that you have a house loan versus a car loan. There's a lot of similarity there".

Table 9 presents an overview of the different possibilities for defining the data domains, as discussed by the interviewees.

Table 9. Approaches to defining data domains.

#	Financial institution type	Data domains	Result summary	# (Interviewees mentioned)
1	Traditional bank	Aligned with the business domains	Data domains should mirror the structure of business domains.	2
		Product-centric	Data domains could be organized around specific product categories, such as mortgages.	2
		Process-driven	Data domains could be defined by processes or customer value streams.	2
2	Challenger bank	Aligned with the business domains	Data domains should be aligned with the business domains (such as financial and sales data) and should not be artificially split.	1
		Process-driven	Data domains are aligned with service lifecycles, such as loan acquisition and loan management processes.	1

#	Financial institution type	Data domains	Result summary	# (Interviewees mentioned)
3	FinTech	Product-centric	Data domains are structured to reflect the operational ownership, with product teams managing and owning the data models.	2
		Aligned with the business domains	Data domains should coincide with business domains, often structured around product offerings.	1
		Process-driven	Domains are organized according to service levels offered, such as loan giving, with similarities grouping different products together.	1

5.3.2 Influencing Factors

The interview participants identified the following factors that influence how data domains can or should be defined: centralization of specific data sets, independence and decoupling of domains, data access, and organizational layout. Similarly to the previous section, the factors are presented first by the type of financial service provider and then ranked by the number of mentions by the interviewees.

Traditional bank interviewees I-02 and I-05 denoted the importance of *centralizing specific data*. Elucidating this, interviewee I-05 stated, "In real life, one size doesn't fit all, and in some cases, you can achieve the independent domains, for example, savings and investments can have their own data sets, but there will always be certain data sets, which should be centralized." The interviewee (I-05) further illustrated it with an example: "For example, if you're talking about the banking industry, the common thing is account. […] in general, development should be product based, but at the same time from a business point of view there should be customer-centric service." Moreover, *independence and decoupling of domains* were mentioned by I-05 and I-09 as aspects that influence the definition of data domains. For instance, I-05 described that the goal from an organizational point of view is "to enable teams to work independently as much as possible" and pointed out that "data should be decoupled, not necessarily decentralized." Interviewee I-09 supported this notion of independence of teams by stating, "People are involved in this, and they have the knowledge. The people responsible for processing data are also responsible for storing the data." Furthermore, interviewee I-05 also underscored *data access* as a potential factor, stating, "For a financial organization, there is quite a big focus on security – who can access data and have evidence […] who has access of data." Interestingly, I-03 presented a distinct perspective regarding the relationship between *organizational layout* and data domains, describing that the two "need to be a bit decoupled otherwise, you will have a mess in the domains with the organization changes as well."

In the interview, a challenger bank interviewee, I-04, also highlighted *domain independence and decoupling* as a factor when it comes to defining data domains. The interviewee noted: "Each team should be able to dictate what data they are saving." Moreover, I-06 discussed the need to *centralize specific data*, stating, "In the end, companies still need to have some central data sets where everything is present." Interviewee I-06 also underscored the significance of *data access*, remarking that not all people "can access everything. Data access is definitely one aspect of that. […] We are not granting access to everybody. There

are special people who can see financial data, and there are even more special people who can see risk data."

Notably, all FinTech interviewees (I-07, I-08, I-10) identified *independence and decoupling of domains* as an influencing factor. For instance, I-10 connected domain independence to microservices, explaining that it makes sense for the teams to take responsibility for the data. Otherwise, you might not have "independent teams [and] then you lose the main point of the microservice architecture." Moreover, I-08 noted, "It should rather be decentralized in this manner, and the ownership, along with those design decisions, should sit within the kind of the owning teams." Interviewee I-08 also described the fundamental impact of *organizational layout* on data domains, attributing the phenomenon to "Conway's law." I-08 articulated, "It's the product and service teams that define their domain, and the data they produce falls and sits within that domain naturally." Concluding that "any kind of architectural design tends to follow the organizational layout and structure" (I-08). Additionally, I-10 highlighted *data access* as a consideration, detailing, "There are privacy rules like PCI data, PII data. There's also business-critical data. […] This is one reason that probably in most cases, you need to ask for access to data because data can have different sensitivity levels."

Table 10 presents a synthesis of the factors identified by interview participants as influencing the definition of data domains.

Table 10. Factors that influence defining data domains.

#	Financial institution type	Influencing factor	Result summary	# (Interviewees mentioned)
1	Traditional bank	Centralization of specific data sets	Certain key data sets, such as accounts, need centralization due to their usage across products.	2
		Independence and decoupling of domains	Aiming for domain autonomy to enable independent team operations, with data structures designed for modularity.	2
		Data access	Ensuring secure and controlled data access is a major factor for financial organizations.	1
		Organizational layout	Flexible data domain structure to prevent the disorder from organizational changes.	1
2	Challenger bank	Independence and decoupling of domains	Teams should manage their data independently for control and effectiveness.	1
		Centralization of specific data sets	Some level of central data sets is necessary to maintain a holistic company view.	1
		Data access	Access control is critical, with only designated personnel able to view sensitive financial or risk data.	1
3	FinTech	Independence and decoupling of domains	Microservices architecture supports data domain independence. Responsibility for data should reside with the owning team.	3
		Organizational layout	The data architecture tends to align with the organizational structure, as described by Conway's law.	1
		Data access	Access to data must consider privacy and sensitivity levels, requiring specific permissions.	1

5.4 Data Mesh

This section presents results for what value would the adoption of a data mesh architecture have for financial service providers (RQ4). Four main areas emerged from the interviews. The first area focused on the benefits that adopting a data mesh architecture could bring. The second area explored the challenges associated with migrating to a data mesh architecture. The third area identified the catalysts for change that would facilitate the shift towards a data mesh architecture for financial service providers. Lastly, the fourth area focused on the value of adopting a data mesh architecture for different financial service providers. Therefore, the results are presented in the same sequence: benefits and challenges of adopting a data mesh architecture, catalysts for change, and value of data mesh.

5.4.1 Benefits

During the interviews, the interviewees identified the following benefits that the adoption of a data mesh architecture might entail: enhanced governance of data, clear responsibility and ownership of data, increased efficiency, increased independence of domains, data democratization, scalability, viewing data as a product, and adaptability across regions. The benefits are grouped according to the type of financial service provider and presented starting from the most frequently mentioned benefit by the interviewees.

The interviewees from traditional banks, I-01, I-02, I-03, and I-05, pointed out *enhanced governance of data* as an advantage. The interviews revealed that data mesh facilitates to "ensure good data quality" (I-02), enhances "visibility of what this data is about" (I-03), and establishes "an agreement on how data should be made available in the company" (I-01). Interviewees I-02, I-03, and I-05 also recognized *clear responsibility and ownership of data* as a benefit. I-02 explained, "When you don't have this logical split, then everyone kind of everyone is confused about who owns what. […] When you have this split […], it is much easier to implement these data management processes and explain the rules." Moreover, I-03 added to this, stating that it brings about "much clearer responsibility."

Furthermore, interviewees I-02, I-05, and I-09 also highlighted *increased efficiency* as a benefit of adopting data mesh architecture. For instance, I-02 mentioned, "The main benefit of this concept is that the process is super efficient, that there are no bottlenecks." Interviewees I-05 and I-09 highlighted it from the perspective of teams being able to utilize domain knowledge. I-05 described, from the perspective of a central data team, "they know data, but they don't know *the* data. They need to come to the same people and ask what the meaning of the data is […] to be able to interpret it." Another benefit mentioned by interviewee I-09 was *increased independence of domains* as teams would be able to "deal with [data problems] localized." The interviewee (I-09) further elaborated that "you can benefit from the decentralized or independent way of managing data. You have isolated teams or domains that move at their own speed and have their own roadmaps." Notably, *data democratization* was pinpointed as a benefit by I-02, underscoring that data mesh supports creating "consumable data products for other platforms and products to use."

Challenger bank interviewee I-04, similar to traditional bank interviewees I-05 and I-09, highlighted that data mesh enables teams to utilize domain knowledge as a way to *increase*

efficiency. "At the moment, the data is in the data warehouse, and I should know everything. People […] ask for very detailed information. Information that I would like to ask the product team, who can tell you where it is, and we can then direct you to the right table. […] We are responsible for this data, but at the same time, we lack the necessary domain knowledge" (I-04). Moreover, interviewee I-06 mentioned *increased independence of domains* as an advantage of data mesh. I-06 pointed out that it empowers "some business area, who is capable and willing to manage their own data in the most effective way."

From the perspective of FinTechs, interviewees I-08 and I-10 discussed that data mesh enables *clear responsibility and ownership of data*. I-08 noted, "The ownership along with responsibilities for design should reside within the owning teams." Both I-08 and I-10 also spoke to the benefit of *enhanced data governance*, with I-08 observing that as data mesh enhances processes surrounding data: "it also improves not only the security posture but the trust nature and aspects there as well. It improves the integrity all around." I-10 supported this view, emphasizing that there is "value to have contractual data. To think about the data you're providing to the warehouse as a contract before publishing it or before it becomes a problem." Uniquely, I-08 highlighted the advantage of *scalability* that data mesh "brings about more scalable management of the whole data ecosystem."

Furthermore, the interviewee (I-08) touched upon *increased efficiency* as a benefit, mentioning that the data ecosystem is "more streamlined and aligned and it's more covering processes as well." Interestingly, I-10 also presented a benefit with regard to the cultural changes that data mesh could bring; outlining that it could enable organizations to *view data as a product*. "This would be something that would maybe help grow the teams into data owners, not only system owners." (I-10). Furthermore, I-07, as the sole interviewee, highlighted the capacity of data mesh to enhance *adaptability across different regions*. "Different countries may have their own business processes. The business needs in different regions can be different. […] there may be completely different regulations. […] If you have to consolidate all those needs into one single place, then this might be too complex" (I-07).

Table 11 illustrates the benefits of implementing a data mesh architecture across different types of financial institutions, as cited by interviewees.

Table 11. Benefits of adopting a data mesh architecture.

#	Financial institution type	Benefit	Result summary	# (Interviewees mentioned)
1	Traditional bank	Enhanced data governance	Decentralization helps to achieve better data quality, improve visibility, and achieve standardization across the company.	4
		Clear responsibility and ownership of data	A logical data structure clarifies ownership and responsibilities.	3
		Increased efficiency	Removing bottlenecks between domain and centralized data teams leads to more efficient processes.	2
		Increased independence of domains	Decentralized management allows domains to operate at their own pace and resolve issues locally.	1
		Data democratization	Data mesh supports creating accessible data products for wider use within the organization.	1

#	Financial institution type	Benefit	Result summary	# (Interviewees mentioned)
2	Challenger bank	Increased efficiency	Domain knowledge enables teams to manage data more effectively.	1
		Increased independence of domains	Different business areas are empowered to manage their data effectively and independently.	1
3	FinTech	Clear responsibility and ownership of data	Teams have clear ownership, promoting accountability for the data they manage.	2
		Enhanced data governance	Data mesh strengthens processes around data, improving security and integrity.	2
		Scalability	Decentralization offers a scalable approach to managing the data ecosystem.	1
		Increased efficiency	A streamlined data ecosystem aligns with processes, enhancing operations.	1
		Viewing data as a product	Teams grow into data ownership, transitioning from system to data product managers.	1
		Adaptability across regions	Data mesh allows for tailored compliance and business processes in different global regions.	1

5.4.2 Challenges

The interviewees outlined seven challenges related to adopting a data mesh architecture: data mesh is not a one-size-fits-all solution, ambiguity in conceptual understanding, changing the culture of an organization, convincing the decision-makers, the efforts required for development, and cross-domain aggregation of data. Challenges are presented firstly for traditional banks, then for challenger banks, and lastly, for FinTechs. Each benefit is ranked in order of the number of mentions by interviewees from the specific type of financial institution.

Among traditional banks, four interviewees (I-02, I-03, I-05, I-09) shared that *development efforts* would be a challenge in transitioning to a data mesh architecture. I-03 observed, “If you start from scratch, then it’s quite natural that you’re building up this common interface on data products. […] However, that will be more of a struggle for this established organization. […] data mesh or the data products require quite some re-development.” I-05 specifically stated that for legacy systems “it’s difficult to refactor anything.” Moreover, interviewees I-01 and I-05 emphasized that data mesh is *not a one-size-fits-all solution*. I-01 described data mesh as “an architectural agreement, not a full solution”, further elaborating that it’s “an idea, but it can be made or taken to life in so many different forms and ways. […] How you do it that’s the key to success.” I-05 reinforced this view by pointing out that it’s “not a tool which will solve all our problems. It doesn’t exist in the world. Data mesh is more about concept. It’s about the way of working […] not about practical technology”.

Interviewees I-03 and I-05 also underscored *changing an organization’s culture* as a challenge. Interviewee I-03 remarked, “The cultural change is the hardest.” I-05 discussed that not many people are used to “being responsible not only for applications but also for data,” highlighting that such a cultural shift requires time and concerted effort: “You need to bring this culture. Teaching spends quite a lot of time to get this culture to work in an

organization. A culture that an organization will adopt." Furthermore, I-03 and I-09 pointed out the challenge of *aggregating data across domains*. For instance, I-03 stressed that "data mesh requires optimal data aggregation capabilities." I-01 noted the *ambiguity in conceptual understanding* of data mesh, likening it to agile methodology: "Because at the end of the day, it is like agile – it's interpreted in so many ways, depending on the country." Additionally, I-03 discussed the challenge of *convincing decision-makers* of the value of data mesh: "You need to sell it to the decision-makers that this is a beneficial activity for the long term. Otherwise, they won't understand the business value why to do that, because, for them, it is a technical activity."

Considering the perspectives of challenger bank interviewees, I-04 and I-06 both identified *cross-domain aggregation of data* as a challenge. I-04 explained, "If there is some regulation from the central bank, you may have to rebuild things to get your piece of data from every domain. This is because different reports can be more complex." Similarly, I-06 expressed concerns about consolidating data from numerous sources, stating, "Nobody wants to start splicing together management reporting from 17 different sources." In addition, I-06 articulated the complexity of *development efforts*, noting, "How long this process drags on and when you reach this point when you have all the pieces in the new place, then you're still maybe stuck five or more years maintaining those parallel systems at the same time." The interviewee (I-06) also recognized the *ambiguity in conceptual understanding* of data mesh, remarking, "To be completely honest, I haven't fully even figured out what it is. The descriptions tend to be a little bit vague."

Furthermore, interviewee I-04 identified organizational *culture change* as a primary obstacle, stating, "The main difficulties are not technical but rather related to people. It's difficult to tell people that now you are a self-organizing team, which they would rather not do." I-04 also highlighted the challenge of *convincing decision-makers* of the value of data mesh. "You would also need to show how much it will bring in some kind of business revenue and justify what does it make better. If I could justify that we are growing and [...] it could increase the quality of the data. Through this, it would be possible to bring about this change" (I-04).

In the interviews with FinTechs, interviewee I-08 also explained that data mesh is *not a one-size-fits-all solution*, requiring a tailored approach instead. The interviewee (I-08) noted, "Nobody does anything by the book. Usually, you take the best pieces that work in your context […] there isn't one size fits all." Additionally, I-08 pointed out the a*mbiguity in the conceptual understanding of data mesh*, describing it as an "abstract concept." Echoing the viewpoints of interviewees I-03 and I-04 from banks, interviewee I-08 likened *cultural changes* to be the most difficult. "Oftentimes, what limits organizations in making changes are the organizational limits or inflexibilities […] this also tends to be a major blocker for some of the traditional financial institutions like banks in doing innovation because that would also require cultural changes." (I-08). The challenge of *development efforts* was shared by I-10. The interviewee noted, "It's not trivial to do any migration. It is a longer, painful process, and you have to have the data in parallel for a while." *Cross-domain data aggregation* was highlighted as another challenge by I-10, stating that "it may be difficult

to correlate the data together between data warehouses." Notably, the interviewee (I-10) also shared that it might be difficult to *convince the decision-makers*: "If a company has already built an existing system and the pipelines are working, maybe it's a tough sell to start pushing this change."

Table 12 provides a summary of the challenges of adopting a data mesh architecture for financial service providers.

Table 12. Challenges of adopting a data mesh architecture.

#	Financial institution type	Challenge	Result summary	# (Interviewees mentioned)
1	Traditional bank	Development efforts	Transitioning to a data mesh architecture requires significant redevelopment, particularly challenging for established organizations with legacy systems.	4
		Not a one-size-fits-all solution	Data mesh is a conceptual approach that must be customized to fit each organization's unique needs.	2
		Changing the culture of an organization	Adopting data mesh requires a shift in culture towards data ownership and responsibility.	2
		Cross-domain aggregation of data	Efficient aggregation of data across domains can be a complex task.	2
		Ambiguity in conceptual understanding	Data mesh is often conceptually misunderstood, likened to the varied interpretations of agile methodology.	1
		Convincing the decision-makers	Demonstrating the long-term business value of data mesh to stakeholders.	1
2	Challenger bank	Cross-domain aggregation of data	Complexities arise when aggregating data from multiple domains for reporting.	2
		Development efforts	Parallel system maintenance during the transition to data mesh is resource-intensive.	1
		Ambiguity in conceptual understanding	The concept of data mesh is perceived as vague and not fully understood.	1
		Changing the culture of an organization	Shifting to self-organizing teams that manage their own data is a significant cultural change.	1
		Convincing the decision-makers	Requires justifications for its potential to enhance business revenue and data quality.	1
3	FinTech	Not a one-size-fits-all solution	Data mesh must be adapted to the specific context of the organization, and not every aspect may be applicable.	1
		Ambiguity in conceptual understanding	Data mesh is an abstract concept, which people interpret differently.	1
		Changing the culture of an organization	Cultural adaptation as the highest barrier to the adoption of a data mesh architecture.	1
		Development efforts	Reorganizing existing data systems is difficult and requires running parallel systems.	1
		Cross-domain aggregation of data	Correlating data across different domains and warehouses is complex.	1
		Convincing the decision-makers	It can be difficult to motivate the change to data mesh when existing systems and pipelines are functioning adequately.	1

5.4.3 Change Enablers

The interview participants identified the following catalysts for adopting a data mesh architecture: strong data governance, visionary leader, centralized data specialists, effective communication for change, introducing appropriate tooling, iterative implementation, and innovative leadership culture. The results are organized initially according to the type of financial service provider and subsequently presented based on the frequency of mentions from the interviewees.

Strong data governance was identified as a change enabler for adopting data mesh by the interviewees from traditional banks (I-01, I-02, I-03, I-05). For instance, I-02 emphasized its criticality, stating that data mesh "could be a huge disaster because everyone would treat data as they would like. That's why the precondition for a successful data mesh is strong data governance so that everyone would understand where his or her responsibility starts and ends." Interviewees I-01 and I-05 emphasized having a *visionary leader* at the forefront of adopting a data mesh architecture. For instance, I-05 suggested: "You should have evangelists or somebody who will drive it." To bring upon this change, I-05 also noted the need to have *central data specialists* to support domain teams: "You need to have data stewards in your organization, who will understand the data and who will document it." Moreover, it was outlined that the *change should be effectively communicated,* with interviewee I-05 highlighting the need to "teach, explain, and sell the idea". Another enabler for change, introduced by I-05, was the notion of *introducing appropriate tooling*. I-05 emphasized: "You need to introduce some tools, processes, and practices […] so that cross-teams can find out who the owner is, how to access it, and what the structures are."

Challenger bank interviewees I-04 and I-06 also emphasized the need to have *data experts in data domains*. Whilst I-04 focused on finding "a very strong data person for each domain," I-06 advocated for "a decent data team for each data product." Interviewee (I-04), however, emphasized that it all starts with *strong data governance*: "First of all, it should start with having a broad-based and strong data governance in place and ensure that every domain starts using it." In addition, I-04 highlighted having a *visionary leader*. The interviewee (I-04) added to previous viewpoints, specifying that it should be "someone with very strong technical knowledge." I-04 further emphasized the importance of *effective communication for* implementing this *change*, stating the need to "communicate things correctly to people. Communication is constantly a cross-domain issue and needs to be coordinated." In addition, it was noted by I-04 that the transition to a data mesh architecture could be done *iteratively*. "The introduction of data mesh could be successful if you do it one fragment at a time" (I-04).

FinTech interviewees (I-07, I-08), similarly to representatives from banks, also identified *strong data governance* as the key prerequisite for change. "There needs to be a governance visibility into the access aspects and other kind of security aspects there as well. […] this needs to be in place first" (I-08). I-07 expanded on this, specifying, "Data governance has to be in place to make sure that we have an understanding of how many domains we have, what is inside of the domain, how to avoid data duplication." Interviewee I-10 underscored the necessity of having *central data specialists* to support the domains, noting "a centralized

data team that helps domains specify those contracts." In addition, I-08 outlined three other drivers that would enable organizations to move towards data mesh. Firstly, similarly to I-04 (challenger bank), I-08 emphasized that it should be implemented *iteratively*: "It might be more efficient to employ a few key practices and principles, not build a full-fledged platform that aligns with all of the principles." Secondly, the requirement for *appropriate tooling* was also emphasized. "If your organization is growing and if you are a heavy data-driven organization […], then self-service also becomes key" (I-08). Lastly, I-08 uniquely stressed the need for an *innovative leadership culture*, noting, "It comes down to leadership. Start from transforming leadership radically and instill a new culture of innovation and agility."

Table 13 presents an overview of change enablers necessary for the adoption of a data mesh architecture, as identified by interviewees.

Table 13. Change enablers for transitioning to a data mesh architecture.

#	Financial institution type	Change enabler	Result summary	# (Interviewees mentioned)
1	Traditional bank	Strong data governance	A foundational requirement to ensure clear responsibility and maintenance of data quality.	4
		Visionary leader	Evangelists or someone to drive the adoption of a data mesh architecture and to instill a new data-centric culture.	2
		Central data specialists	Specialists to support domain teams with data stewardship and documentation.	1
		Effective communication for change	Communicating the benefits and processes of data mesh for successful adoption.	1
		Introducing appropriate tooling	Tools and practices for managing metadata and data access.	1
2	Challenger bank	Expertise in data domains	Data domains need to have a data expert(s) who would oversee the management and structuring of data.	2
		Strong data governance	A broad-based and strong governance framework is necessary to ensure that each domain uses data correctly.	1
		Visionary leader	A technically knowledgeable leader to guide the transition.	1
		Effective communication for change	Proper communication strategies are necessary for coordinating the change.	1
		Iterative implementation	Adoption should be gradual, focusing on small, manageable changes over time.	1
3	FinTech	Strong data governance	Clarity in governance to manage access and security before transitioning to data mesh.	2
		Central data specialists	Specialized central teams to assist domain teams in defining and managing data contracts.	1
		Iterative implementation	Adopting a data mesh architecture in smaller, manageable steps is more pragmatic and likely to be successful.	1
		Introducing appropriate tooling	The right tools are critical for self-service and scalable management of the data ecosystem.	1
		Innovative leadership culture	Leadership must be open to innovation and foster a culture that supports agility and change.	1

5.4.4 Value

The interviews revealed that whilst some interviewees recognized that data mesh contributes new value, others found it to have limited value in their current context. These findings are organized by the type of financial service provider and ranked by the number of times interviewees mentioned each value aspect.

Significantly, all traditional bank interviewees (I-01, I-02, I-03, I-05, I-09) found data mesh to be *valuable*. One interviewee noted, "There is a huge benefit and opportunity to use this concept, but only when some preconditions or prerequisites are fulfilled. […] it could bring new value. […] It boosts business and delivers added value for customers." (I-02). Noteworthy is that two interviewees from the same bank (I-01 and I-05) indicated they are progressing towards data mesh. I-05 shared this experience: "We started this data mesh initiative two weeks ago. […] It takes time to sell it in a bigger organization." Interestingly, I-09 mentioned an implicit alignment with the adaption of the principles, "the ideas we use, we don't call it [data mesh], and none of our internal documents mention the data mesh concept. But if you look at the concept from a more abstract level, then it's something we've been moving towards." The interviewees also touched on the *value* of data mesh being *dependent on the organization*, specifically, on the "maturity of an organization" (I-02), "size" of an organization as having data centralized has a "huge advantage to a certain size" (I-05), and the number of "independent product areas" (I-09).

Contrastingly, challenger bank interviewees (I-04, I-06) found that data mesh would have *limited value in their current context*. I-04 remarked, "In our current context, it probably wouldn't have much value. Our products or domains are small enough, and the amount of data is still small enough to manage it in one place. If the amount of data becomes so large that the data warehouse should be split, then something like this should be thought of." From a resource perspective, I-06 added that "it seems to be a luxury for bigger organizations, who can have those meaningful teams around those individual data products." The interviewees also explicitly underscored that the value of data mesh depends on the organization; however, they did not offer insights that differed from those provided by traditional bank interviewees.

From the perspective of FinTechs, I-07, and I-10 also did *not recognize data mesh to be very valuable* in their current context. For instance, I-07 elaborated, "The simplest answer is that we are too small for this kind of thing. [...] It's about the number of dependencies and how many resources you have. [...] If we cannot manage it in a central data warehouse anymore. and the system is going to be already that big, then yes, this is going to make sense. But at least in our case, then this is going to be a bit pointless overhead for us." All FinTech interviewees (I-07, I-08, I-10) discussed the importance of the organization's specific context in determining the value of data mesh. I-08 noted, "It depends on the context. It might not be the right time for or not the right thing at all for an organization to aspire towards data mesh. There might not be the right time for it." Interestingly, the interviewee (I-08) also pointed out "future growth prospects" as a consideration, adding to factors identified by traditional banks in determining the value of data mesh for an organization.

Table 14 illustrates the value of adopting a data mesh architecture as perceived by traditional and challenger banks as well as FinTechs.

Table 14. Value of adopting a data mesh architecture.

#	Financial institution type	Value	Result summary	# (Interviewees mentioned)
1	Traditional bank	Brings new value	Seen to bring value for its potential to enhance data access and support business decision-making.	5
2	Challenger bank	Limited value in the current context	Perceived as less valuable in smaller or less complex settings where data can be effectively managed centrally.	2
3	FinTech	Limited value in the current context	Considered not immediately beneficial and is seen to bring unnecessary overhead.	2

6 Discussion

This section discusses the findings, answers the research questions, outlines the limitations of the thesis, and gives proposals for future research.

6.1 Current State of Data Management

In approaching RQ1 on how data is managed by financial service providers, the goal was to gain a deeper understanding of the current data management practices across traditional banks, challenger banks, and FinTechs. In particular, the aim was to understand how data is managed. The main findings suggest that data management can be considered to be somewhat dependent on the type of financial service provider. Although similarities exist between all three types of financial service providers.

The research indicates that both traditional banks and challenger banks rely on legacy monolithic source systems, designed originally for on-premise operation, that generate data. Data from these systems is then typically managed through operational databases or database platforms and consolidated into one or many centralized repositories for analytical purposes, including regulatory or management reporting. Interviewees discussed using many centralized data warehouses and data lakes. The findings are consistent with what is being described in the literature [5], [16], [17], [37], [39], that highlight how centralized structures often fail to adequately meet business demands. However, in comparison to the interview findings, [5] and [16] specifically outline that, as the production and consumption of analytical data become more decentralized, the centralized structures do not match the logical flow of data within the organization.

The findings also indicate that data delivery to these centralized repositories commonly occurs through offline delivery mechanisms such as ETL pipelines, which may be considered the prevalent delivery mechanism. For centralized data architectures, these ETL pipelines are often also centrally managed [5], [31]. The reliance on legacy source systems, centralized data architectures, and ETL processes among banks could be attributed to different factors. Firstly, these systems have proven to be highly reliable over time, which is crucial for banks that prioritize stability in their operations. Additionally, the use of these traditional systems may reflect a cautious approach to change, which is influenced by the regulatory compliance and extensive reporting requirements that banks must meet. This could enable banks to maintain control over their data processes, minimizing risks associated with newer, less-tested technologies.

However, the data reveals subtle distinctions that set the two types of banks apart in their data management approaches. The data implies that challenger banks only depend on offline data delivery systems, such as ETL pipelines. This could indicate either the lack of a real-time data processing requirement or a limitation in their capacity to adopt such technologies. On the other hand, traditional banks appear to also utilize, to some degree, near real-time delivery mechanisms, such as Kafka. This may be a response to the evolving landscape of the financial sector, including increased competition from FinTechs or increased customer demand for instant and on-demand services. As revealed by the interviews, challenger banks tend to also have a single team responsible for all aspects of data management without

dedicated personnel specifically for data governance or data quality – roles that may be commonly found in traditional banks. As challenger banks have a smaller scale and fewer product offerings, their data systems could reflect this, potentially carrying less complexity than traditional banks. This could further be supported by data showing that challenger banks face fewer challenges than traditional banks. Therefore, one might argue that the primary distinction in how data is managed in challenger banks and traditional banks lies in their scale.

The findings indicate that data management in FinTechs differs from banks in a number of ways. Firstly, the interviews suggested that FinTechs utilize a microservice-based architecture. As this also brings about decentralized operational data management, with each microservice handling its own database, it implies that the teams responsible for the microservices are also more connected with the data they handle. Secondly, the interviews also suggested a reliance on real-time data processing and the utilization of newer technologies and managed data platforms. While ETL pipelines are still in use, their application seems to be more about suitability for particular use cases rather than being the only option available. However, analytical data management resembles that of banking institutions, as it typically involves consolidating data into central warehouses or lakes for reporting purposes, as underscored in the interviews. Interestingly, this contrasts with insights from a related work [20], which suggests that such centralized structures are incompatible with the dynamic nature of agile development practices that require ongoing updates to applications and data systems.

In summary, operational data management approaches diverge among traditional banks, challenger banks, and FinTechs. Banks have their methodologies rooted in legacy monolithic services. The difference between traditional and challenger banks mainly lies in the scale and complexity of their operations. FinTechs differentiate from both, being inclined towards more modern practices, such as using a microservices architecture. Yet, despite their operational differences, centralizing analytical data, using data warehouses and data lakes, is common across all financial service providers. Findings indicate that the primary difference in analytical data management between traditional banks and challenger banks or FinTechs is that traditional banks maintain a more extensive array of data warehouses and lakes, alongside a larger team of specialized data personnel. This could also be attributed to their wider range of operations. Nevertheless, the use cases for this analytical data – such as regulatory and management reporting – remain consistent across all types of financial institutions.

6.2 Impact of Data Management

The second research question, RQ2, focused on answering how the current data management practices limit financial service providers in achieving their business needs and objectives. The aim was to identify what are the challenges related to data management and understand if and in what ways these impede financial service providers.

The results indicate a correlation that the more established a financial organization is, with a broader range of product offerings, the greater the number of ways in which data

management challenges can impede its business objectives. However, the findings suggest that this correlation does not extend to the business needs since they are externally mandated, necessitating that these needs are prioritized and met above all else.

Interviews with traditional banks identified the most impediments stemming from data management challenges that might affect meeting their business objectives. Amongst these, delayed time to market and additional resource allocation stood out as the most significant impediments. For instance, results indicate that delayed time to market could potentially arise due to challenges such as cardinality and complexity of systems or hierarchical chain of command, which in turn could impact the objective of enhancing customer experience (I-05). Notably, delays in time to market is also mentioned by a related work [37], however, this issue is mainly attributed to unclear data ownership that causes slow data access and poor data quality.

Other impediments mentioned by traditional banks were loss in competitiveness, centralized data team being a bottleneck, and the inability to achieve business objectives altogether. The issue with bottlenecks is also highlighted in related works, which suggest that it arises as organizations scale up the number of data sources [5], [38] and expand their data consumption use cases [5], [16], [38]. The results imply that a lack of real-time data could make it so that financial organizations are unable to achieve their business objectives (such as making time-critical personalized offers) and, thereby, potentially also impede enhancing customer experience and improving financial performance (I-01). Thus, as businesses cannot make data-driven decisions, the organization's ability to stay competitive is impacted [6]. Moreover, the findings also suggest that unclear data ownership could lead to inefficient resource allocation, as the centralized data team may face constant back-and-forth interactions with domain teams, potentially rendering businesses less able to create process efficiency (I-03). Therefore, the results suggest that traditional banks deal with a wider array of impediments that stem from legacy-related and cultural challenges, which could be attributed to their larger scale and lengthier operational history. Although they appear to have more resources available to address these issues, it seems that the variety of impediments surpasses their ability to effectively address them and gives a strong indication of the need for modernizing the data architecture and systems.

Interestingly, however, one traditional bank interviewee (I-02) did not find the challenges to be impediments to achieving business objectives. The same interviewee described having decentralized data ownership, which might imply that decentralization may lessen the impact of data management issues on achieving business objectives by enabling more efficient and autonomous decision-making within distinct business segments.

For challenger banks, when it came to discussing how data management challenges may impede business objectives, the findings revealed only one impediment. A centralized data team was noted as a potential bottleneck, which may be caused by increased maintenance efforts and may cause businesses to be less or unable to create process efficiency (I-06). This is also supported by a related work [32], which highlights that traditional centralized data systems may become inefficient as data volume and diversity grows. As these professionals can largely be consumed by maintenance tasks, there might be little

opportunity to enhance processes or architect new systems. However, the results suggest that whilst there are different challenges, such as issues with data governance and increased maintenance efforts due to legacy systems, they were mostly not seen as impediments to achieving business objectives. This could indicate that whilst challenger banks do not necessarily have a shorter history of operation than traditional banks, they may experience fewer impediments from data management issues from being more specialized, with a narrower range of products.

Notably, the findings suggest that FinTechs do not experience any impediments arising from the data management challenges. This seems to be correlative to the mere two challenges related to compliance and data governance that were outlined in the interviews. Having fewer challenges and no impediments from these challenges may be indicative of their relatively recent establishment in the financial sector, where their technology is more modern and not yet considered legacy. Furthermore, their use of microservices architecture and their current growth phase potentially afford FinTechs agility that might not be present in more traditional institutions, which could help them adapt and overcome challenges more easily.

On the other hand, as business needs (compliance and risk management) are mandated and externally imposed, the findings indicate that financial institutions fulfill them above all else. This means that while data management challenges may complicate the process, they are unlikely to affect whether these needs are met. Nearly all interviewees identified compliance as a business need, with traditional banks explicitly outlining it as their top priority. Notably, the results suggest for traditional banks that the significant allocation of development capacity to compliance-related activities can divert resources away from other important data management activities, potentially impacting the achievement of business objectives (I-05). One might argue, however, that the complexity of regulations for traditional banks could be magnified by the cardinality and complexity of legacy systems, with this dynamic creating a cycle where past regulations have shaped complex system designs, which now complicate adhering to new regulatory changes. Interestingly, while complying with regulations was mentioned, it did not emerge from the findings as a significant concern for challenger banks or FinTechs. This may suggest that less complex systems or more modern systems facilitate easier compliance with regulatory changes and also lends weight to the argument that it is the complexity of systems in traditional banks rather than the regulations themselves that intensify compliance difficulties.

To summarize, the research findings indicate a correlation between the longer establishment and product diversity of financial service providers and how data management challenges impact achieving their business objectives. Traditional banks, with broader product offerings and older systems, face numerous impediments like delayed time to market and resource misallocation, which hinder their ability to enhance customer experiences and maintain competitiveness. In contrast, challenger banks experience fewer impediments despite similar challenges, likely due to their narrower product focus. FinTechs, benefitting from modern technologies and agile structures, report no impediments from data management challenges. Despite this, all institutions emphasized the importance of meeting

externally mandated business needs, such as regulatory compliance. As business needs take precedence over business objectives, meeting these needs remains unaffected by data management challenges.

6.3 Possibilities for Defining the Data Domains

The third research question, RQ3, explored how data domains could be defined for financial service providers to support the fulfillment of their business needs and objectives. The goal was to understand what could be an appropriate approach to organize data into domains.

The interviews suggested that the data domains could be defined by products or processes and could be in alignment with the business domains. Yet, from the findings, no single approach distinctly stood out as the most advantageous.

From the interviews, the most commonly proposed approaches included organizing data by products or processes. In a product-oriented approach, data could be structured around specific financial products like mortgages, credit cards, and savings accounts. This approach groups all related data, possibly making it more straightforward to manage and analyze. This approach could benefit from the stability of financial products (I-02), potentially offering a consistent approach for defining data domains. On the other hand, a process-driven approach involves organizing data according to the various stages of a customer's interaction with the financial institution, from account opening to loan approval and ongoing loan management (I-04). This method emphasizes the flow of operations, ensuring that findings support the execution of business processes. Although both methods have distinct advantages, the preferred approach was not clear, indicating that both could work well depending on the organization's specific needs and perhaps suggesting that the most important aspect could be the consistency of how the domains are defined [33].

Additionally, the interviews highlighted that data domains should align with business domains, influenced by the organizational structure of the business. This could be considered particularly pertinent in large organizations like banks, where products or processes and their data span multiple divisions. By first organizing data by business domains and subsequently by products or processes, organizations can ensure that data remains closely linked to its appropriate domain. In contrast to the interviews, whilst the related works [5], [17], [31], [37], [41], [42], [47], [50], [55] supported the same approach, they did not subdivide the data domains further than the business domains. For example, a related work [55] organized data domains strictly according to business domains but included multiple data products within these domains, each corresponding to different financial products. This discrepancy suggests that the interviewees might have conflated these data products with data domains, or that distinct offerings of a product, such as car and student loans, are treated as separate data products within the same domain. In this case, an important consideration to make is that the categorizations, as noted by I-10, are quite similar, which might result in artificially fragmented data.

Nevertheless, an important consideration was highlighted in one interview: aligning data too closely with business domains can lead to issues in the future (I-03). Business domains tend to change more frequently than the underlying data used in the financial sector,

meaning that changes to the organizational structure could necessitate changes in data domains. This may indicate that while aligning data with business domains may enable domain independence and faster time to market (I-05), it could risk data disorganization as business structures can be impersistent and evolve. For example, merging divisions that offer the same product could lead to redundant data sources. Hence, one might argue that whilst the literature suggests alignment with business domains, it does not consider the industry-specific needs, such as the data retention periods in the financial sector, where data often outlasts business model changes.

However, this does not necessarily imply that data domains cannot align with business domains; rather, it seems that shaping these domains should be in alignment with business vision and strategy. This alignment between business objectives and data mesh was highlighted also by [5]. The involvement of business stakeholders in defining data domains is important as without it, there could be the risk their strategies may not be fully supported by the underlying data structures. Whether it is a FinTech company strategizing for geographic expansion, a challenger bank broadening its product range, or a traditional bank focusing on customer experience in a stable market, data domains should be structured to support the business strategies, recognizing that such business objectives are usually not something that would abruptly change. Therefore, it does not seem feasible to segregate the data functions from the broader business objectives, but rather the task of defining data domains extends beyond the architects alone and requires a collaborative effort with business stakeholders.

On the other hand, though interviewees agreed on the importance of keeping product or process-related functionality in its data domain, there was an unanimity that some data sets, such as customer or account data, should be centralized. A related work [6] classifies this data into an aggregate data domain as the data is aggregated from different data products from other domains to facilitate the intersection of customers and products, allowing for a holistic view of the customer.

In conclusion, there seems to be no clear way how to define the data domains. The findings from the interviews point towards a product or process-oriented approach that aligns with business domains as well as highlight that certain data sets should be centralized. However, in choosing how to do it, it needs to be adapted to the specific context and needs of the organization, taking into consideration and aligning the business objectives with the data strategy and consequently, understand how persistent the domains should be and how would organizational changes impact the data domains.

6.4 Value of Data Mesh

In addressing the fourth research question (RQ4), the value of the adoption of a data mesh architecture on financial service providers was explored. The intent was to understand if data mesh could have value in terms of addressing the impediments faced by financial organizations.

The findings indicate that although the value of data mesh depends on an organization, there seems to be a strong correlation between the maturity and scale of an organization and the

value derived from implementing a data mesh. Specifically, traditional banks with larger scale and more maturity found data mesh to be highly valuable, whereas challenger banks and FinTechs in their current contexts seem to be deriving less benefit.

The findings revealed that traditional banks could find significant value in adopting a data mesh architecture. This could be attributed to their maturity and extensive range of product offerings, which seem particularly suited to benefit from distributed data architecture and teams [34]. The benefits seemed to mainly revolve around decentralizing data ownership as it could enhance data governance, clarify responsibilities and ownership of data, and increase the efficiency and independence of domains. Notably, these benefits could address impediments faced by traditional banks, such as delays in time to market, decreased competitiveness, and bottlenecks created by centralized data teams interacting with domain-specific teams. However, the findings indicate that adopting a data mesh architecture poses challenges, particularly from the development and cultural perspective. The findings suggest development efforts to be significant for traditional banks due to their reliance on complex legacy systems and the specific requirements of data migration in the financial sector, such as complying with data retention laws and maintaining long-term contracts that may span decades. Consequently, data migration might necessitate maintaining parallel operations to meet regulatory standards and fulfill contractual obligations. Notably, however, cultural change was noted as the hardest, likely due to their long establishment, which might make the organizations more resistant to change. Notably, culture-related changes were also described as the most difficult by a related work that discusses data mesh in the financial sector [18]. Nonetheless, the potential advantages of a data mesh might outweigh these obstacles, as it seems to address the impediments that traditional banks are facing. Moreover, banks seem to put significant focus on data governance (supported also by [18]), considered as one of the key catalysts for change, and have resources – including visionaries and data specialists – to drive and implement this change.

Conversely, interviews with challenger banks and FinTechs indicate that the adoption of a data mesh architecture is of limited value in their specific contexts due to their smaller scale. This is similar to existing works [32], [34], where it is suggested that data mesh should not be used for small-scale data operations. For challenger banks, data mesh appears to be less beneficial as their current data management practices do not significantly impede meeting business objectives. Centralized data teams acting as bottleneck was noted as the sole impediment, which this architecture could potentially address by decentralizing data ownership. Despite this, the effort required for adoption, particularly given their reliance on legacy systems and the complexities of data migration, might not be justifiable (discussed in [32]). FinTechs, however, unanimously did not perceive data mesh to be beneficial in their current context. Whilst they did not report any impediments resulting from their current data management practices, it could be important to consider their focus on future growth (I-08). Implementing a data mesh architecture early in their development could prove highly advantageous, helping to preemptively reduce data migration challenges as they scale. Moreover, the transition to a data mesh is potentially easier for FinTechs, as they already possess a domain-driven mindset (mentioned also in [34]) and typically utilize newer technologies, which can facilitate the adoption process.

Notably, the interviews revealed an interesting viewpoint where organizations might move towards data mesh, by employing one or more principles, without formally recognizing it as data mesh (I-09). The findings suggest that the organizations might implement elements of the architecture simply because "it just works" (I-08). This observation suggests that full implementation of a data mesh may not be necessary for organizations to derive value from its principles. Instead, an iterative approach towards adopting data mesh might prove beneficial. For traditional banks, focusing on achieving quick, cost-effective wins could be considered especially important as it could help justify the transition for expansive legacy systems [16]. Supporting the iterative approach, findings from an interview with one traditional bank revealed the decentralization of analytical data ownership with strong data governance, aligning with the principles of data mesh (I-02). The absence of any reported impediments to meeting business objectives in this case further implies that even partial adoption of data mesh principles can significantly benefit financial organizations.

To conclude, the value of data mesh seems to depend on the organizational maturity, scale, and growth prospects. Findings indicate that data mesh could be very valuable for larger and more established banks due to their maturity and variety of product offerings. The benefits of data mesh could help these organizations address issues such as slow market response and competitiveness, albeit they come with challenges related to the development efforts of legacy systems and cultural adaptation. On the other hand, challenger banks and FinTechs see less immediate value, as their current practices do not heavily or not at all hinder their objectives. However, FinTechs could find early adoption beneficial for future scaling. Additionally, the research uncovered that some financial institutions are moving towards a data mesh by integrating its principles into their operations without labeling it as such, suggesting that data mesh could be valuable even if the full architecture is not adopted.

6.5 Limitations

The research employed semi-structured interviews as the only data collection method, which could be a limitation because the findings are subject to the personal biases and experiences of the interviewees, possibly not capturing all relevant viewpoints. Nevertheless, it enabled the author of this thesis to gain a deep understanding of expert views related to data management, which was the objective of this thesis. Accordingly, this approach allowed for an in-depth exploration of the subject and provided the opportunity to ask follow-up questions to enrich the data collected.

A possible limitation of the thesis was the presence of a single interviewer, which could introduce biases and preconceptions that might influence the data collection and analysis processes, as well as the risk of misinterpreting the interview responses. To avoid these concerns, the researcher actively questioned and examined their own biases and assumptions throughout the study. All interviews were recorded and transcribed, allowing the researcher to revisit and review the recordings for clarification if necessary. Additionally, the interview guide was structured to have different questions for each research question, enabling the exploration of answers from multiple perspectives.

Moreover, the research for this thesis was limited to interviewees from the Baltic region. As a result, the findings may not be fully reflective or applicable to financial institutions in other regions that may operate under different regulatory, economic, or cultural environments. To counteract this limitation, the study included specialists with extensive experience in various companies, some of which operate beyond the Baltic region, aiming to reflect the broader financial services sector better.

A potential limitation is also that the interviews were conducted with a limited number of professionals from specific financial institutions. Hence, the findings may not accurately reflect the financial services sector in the Baltics as a whole. Similarly to the previous limitation, a diverse sample of interviewees was selected to avoid this. This was done by including interviewees with extensive tenures and experiences across different companies and roles. For instance, the sample included roles such as head of data quality, head of analytics, IT chief architect, and senior data infrastructure lead.

Another limitation of the study is that interviewees may not have fully understood or could have misinterpreted the interview questions. Multiple questions were asked on the same topic to gain insights from multiple perspectives to mitigate this risk. Additionally, interview participants were encouraged to express any concerns or questions to avoid misunderstandings and overlooking important areas. For instance, the interview guide included the following question to encourage participants to discuss any additional relevant information: “Before we conclude, would you like to share any additional insights or topics not yet discussed?”

The novelty and abstract nature of the data mesh concept is also a possible limitation, as there was a risk of different interpretations of the concept by the interviewees. To avoid this, the definition and principles of data mesh were clarified with interviewees as necessary. Additionally, follow-up questions were employed whenever there was uncertainty about whether interviewees shared a consistent understanding of the concept to ensure a consistent understanding of the subject matter.

Additionally, the inability or unwillingness of participants to share critical or sensitive information due to non-disclosure agreements or confidentiality commitments presents a potential limitation to the research. Interviewees might have restrictions placed by their employers and partners to prevent the disclosure of company-specific information during interviews. This can potentially impact the comprehensiveness of the research findings.

6.6 Future Research

Given the relatively small sample size of the study, it is recommended that further research involve a larger group of participants from the financial sector. This expanded research could aim to include a broader range of financial institutions, potentially extending beyond the Baltic region, to provide more comprehensive insights, confirming or challenging the findings of this thesis.

Another recommendation for further research, inspired by insights from interviewee I-08, is to do a comparative study across industries, such as healthcare and telecommunications. This would enable to determine whether the findings of this thesis are unique to the financial

sector or share similarities with those in other regulated sectors that handle sensitive data and are subject to regulations.

Additionally, in-depth case studies could be conducted with financial service providers at different stages of adopting a data mesh architecture. Such studies could involve a financial institution that has just begun implementing a data mesh architecture, one in the middle of the transition, and another that has completed the process. These case studies would provide valuable insights into the specific challenges and benefits encountered at each stage of the adoption process, offering a clearer picture of the practical implications and potential outcomes of adopting such an architecture within the financial sector.

7 Conclusion

This thesis aimed to provide an in-depth understanding of how data is managed by financial service providers, including the current practices and impediments of data management. It also sought to deepen the understanding of how data mesh can be defined within the financial sector, focusing on how the data domains can be defined and assessing the potential value that adopting a data mesh architecture might bring. To achieve this, semi-structured interviews with ten industry specialists were conducted. The interview transcripts were analyzed using thematic analysis, which enabled to systematically find key themes from the interviews. As a result, nine themes and fourteen subthemes were identified.

In understanding how data is managed by financial service providers (RQ1), the findings revealed that operational data management differs among financial service providers. Banks often depend on legacy systems, with challenger banks primarily differing from traditional banks in their operational scale. On the other hand, FinTechs favor modern approaches such as microservices. Despite these variations in operational data management, it is common among all financial service providers to centralize analytical data using data warehouses and lakes. The research findings also indicated that data management practices affect financial service providers differently based on their establishment and product diversity (RQ2). More established traditional banks face various impediments such as delayed time to market. Challenger banks and FinTechs, both with narrower product ranges and FinTechs also with modern technologies, experience fewer or no significant impediments.

The research also suggested defining data domains for financial service providers (RQ3) either by product or process, aligning them with business domains, and centralizing certain data sets. However, the specific definition of data domains should be tailored to the organization's context and needs, ensuring alignment between the data strategy and business objectives, while considering the impact of organizational changes. It suggested that the adoption of a data mesh architecture (RQ4) could be particularly valuable for larger, established banks due to their extensive product offerings. However, while challenger banks and FinTechs may see less immediate value, early adoption could benefit FinTechs in future scaling. Additionally, the findings indicated that even the adoption of a few principles of data mesh, such as decentralized data ownership and governance, instead of the full solution, could be beneficial for financial organizations.

Nevertheless, the scope of this research was limited to the financial sector in the Baltic region, which may not fully translate to other regions and industries with different regulatory environments and technological setups. Moreover, while the interviews provided valuable insights, they reflect subjective experiences and interpretations that may not be representative of the sector as a whole.

Future research could extend these findings by involving participants from diverse financial institutions across and beyond the Baltic region to further validate or contest the findings of this thesis. Moreover, conducting in-depth case studies of financial institutions at various stages of data mesh implementation – beginning, mid-transition, and completion – would

provide deeper insights into the specific challenges and benefits of each phase, enhancing understanding of its practical impact within the financial services sector.

Acknowledgments

The author of this thesis is thankful to all the interview participants for their time and for sharing their valuable insights and experiences related to data within the financial services sector.

References

[1] D. W. Enstrom, “Term Definition: Business Domain.” Accessed: Mar. 31, 2024. [Online]. Available: https://www.unified-am.com/UAM/UAM/guidances/termdefinitions/uam_business_domain_94DAA51B.html

[2] Oxford University Press, “Challenger bank.” Accessed: Mar. 31, 2024. [Online]. Available: https://web.archive.org/web/20170818172736/https://en.oxforddictionaries.com/definition/challenger_bank

[3] Atlan, “Data Governance Pillars: Empowering Your Data-Driven Journey.” Accessed: Mar. 31, 2024. [Online]. Available: https://atlan.com/data-governance-pillars/#what-are-the-pillars-of-data-governance

[4] Snowflake Inc., “What is a Data Lake?” Accessed: Mar. 31, 2024. [Online]. Available: https://www.snowflake.com/guides/what-data-lake/

[5] Z. Dehghani, *Data Mesh*, First Edition. O’Reilly Media, Inc., 2022.

[6] J. Christ, L. Visengeriyeva, and S. Harrer, “Data Mesh Architecture.” Accessed: Dec. 10, 2023. [Online]. Available: https://www.datamesh-architecture.com/

[7] Snowflake Inc., “Data Warehouse.” Accessed: Mar. 31, 2024. [Online]. Available: https://www.snowflake.com/data-cloud-glossary/data-warehousing/

[8] E. Evans, *Domain-driven design: tackling complexity in the heart of software*. Addison-Wesley Professional, 2003.

[9] Databricks Inc., “Extract Transform Load (ETL).” Accessed: Apr. 01, 2024. [Online]. Available: https://www.databricks.com/glossary/extract-transform-load

[10] S. Ross, “How the Financial Services Sector Differs From Banks.” Accessed: Jan. 23, 2024. [Online]. Available: https://www.investopedia.com/ask/answers/030415/what-distinguishes-financial-services-sector-banks.asp

[11] J. Kagan, “Financial Technology (Fintech): Its Uses and Impact on Our Lives,” Investopedia. Accessed: Jan. 23, 2024. [Online]. Available: https://www.investopedia.com/terms/f/fintech.asp

[12] J. Chen, “Nonbank Financial Institutions: What They Are and How They Work.” Accessed: Jan. 23, 2024. [Online]. Available: https://www.investopedia.com/terms/n/nbfcs.asp

[13] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qual Res Psychol*, vol. 3, no. 2, pp. 77–101, 2006, doi: 10.1191/1478088706QP063OA.

[14] N. Alam, “Digital Banks Vs Challenger Banks, NeoBanks Vs Open Banks- An Explainer.” Accessed: Jan. 23, 2024. [Online]. Available: https://medium.com/@nafisalam/digital-banks-vs-challenger-banks-neobanks-vs-open-banks-an-explainer-402ece7aac6

[15] B. Marr, *Data strategy: How to Profit from a World of Big Data, Analytics and the Internet of Things*. Kogan Page Publishers, 2017.

[16] J. Bode, N. Kühl, D. Kreuzberger, S. Hirschl, and C. Holtmann, “Towards Avoiding the Data Mess: Industry Insights from Data Mesh Implementations,” Feb. 2023, doi: 10.48550/arXiv.2302.01713.

[17] I. Araújo Machado, C. Costa, and M. Y. Santos, “Data Mesh: Concepts and Principles of a Paradigm Shift in Data Architectures,” *Procedia Comput Sci*, vol. 196, pp. 263–271, Jan. 2022, doi: 10.1016/J.PROCS.2021.12.013.

[18] D. Joshi, S. Pratik, and M. P. Rao, “Data Governance in Data Mesh Infrastructures: The Saxo Bank Case Study,” in *Proceedings of The 21st International Conference on Electronic Business*, Dec. 2021, pp. 599–604.

[19] I. Araújo Machado, C. Costa, and M. Y. Santos, “Advancing Data Architectures with Data Mesh Implementations,” in *Intelligent Information Systems. CAiSE 2022. Lecture Notes in Business Information Processing*, vol. 452, J. De Weerdt and A. Polyvyanyy, Eds., Cham: Springer International Publishing, 2022, pp. 10–18. doi: 10.1007/978-3-031-07481-3_2.

[20] K. Vestues, G. K. Hanssen, M. Mikalsen, T. A. Buan, and K. Conboy, “Agile Data Management in NAV: A Case Study,” in *Agile Processes in Software Engineering and Extreme Programming. XP 2022. Lecture Notes in Business Information Processing*, vol. 445, V. Stray, K.-J. Stol, M. Paasivaara, and P. Kruchten, Eds., Cham: Springer International Publishing, 2022, pp. 220–235. doi: 10.1007/978-3-031-08169-9_14.

[21] S. Bulpin, “Data Mesh 101: Why Federated Data Governance Is the Secret Sauce of Data Innovation.” Accessed: Apr. 08, 2024. [Online]. Available: https://www.mesh-ai.com/blog-posts/data-mesh-101-federated-data-governance

[22] “Introduction to Data Mesh with Zhamak Dehghani.” Accessed: Apr. 02, 2024. [Online]. Available: https://www.youtube.com/watch?v=3Q_XbPmICPg

[23] J. Caserta, J.-B. Dubois, M. Roggendorf, M. Roth, and N. Srinidhi, “Demystifying data mesh,” 2023. Accessed: Apr. 22, 2024. [Online]. Available: https://www.mckinsey.com/capabilities/quantumblack/our-insights/demystifying-data-mesh

[24] W. Kenton, “Financial Sector: Definition, Examples, Importance to Economy.” Accessed: Jan. 21, 2024. [Online]. Available: https://www.investopedia.com/terms/f/financial_sector.asp

[25] A. Barone, “How Banking Works, Types of Banks, and How To Choose the Best Bank for You.” Accessed: Jan. 23, 2024. [Online]. Available: https://www.investopedia.com/terms/b/bank.asp

[26] ClearTax, “Financial Sector.” Accessed: Jan. 23, 2024. [Online]. Available: https://cleartax.in/glossary/financial-sector/

[27] The World Bank Group, “Nonbank Financial Institution.” Accessed: Jan. 23, 2024. [Online]. Available: https://www.worldbank.org/en/publication/gfdr/gfdr-2016/background/nonbank-financial-institution

[28] K. Levitt, “AI Takes Center Stage: Survey Reveals Financial Industry’s Top Trends for 2024.” Accessed: Jan. 23, 2024. [Online]. Available: https://blogs.nvidia.com/blog/ai-in-financial-services-survey-2024/

[29] Z. Dehghani, “How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh.” Accessed: Jan. 25, 2024. [Online]. Available: https://martinfowler.com/articles/data-monolith-to-mesh.html
[30] T. Priebe, S. Neumaier, and S. Markus, “Finding Your Way Through the Jungle of Big Data Architectures,” in *2021 IEEE International Conference on Big Data (Big Data)*, IEEE, 2021, pp. 5994–5996. doi: 10.1109/BIGDATA52589.2021.9671862.
[31] A. Ashraf, A. Hassan, and H. Mahdi, “Key Lessons from Microservices for Data Mesh Adoption,” in *2023 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, IEEE, 2023, pp. 21–28. doi: 10.1109/MIUCC58832.2023.10278300.
[32] A. Dibouliya and V. Jotwani, “Review on Data Mesh Architecture and its Impact,” *Journal of Harbin Engineering University*, vol. 44, no. 7, Jul. 2023.
[33] S. Hokkanen, “Utilization of Data Mesh Framework as a Part of Organization’s Data Management,” 2021.
[34] O. Vinnikainen, “Data Mesh: A Holistic Examination of Its Principles, Practices, and Potential,” 2023.
[35] Zalando, “ZALANDO.” Accessed: Jan. 25, 2024. [Online]. Available: https://www.zalando.com/
[36] Netflix Inc., “Netflix – Watch TV Programmes Online, Watch Films Online.” Accessed: Jan. 25, 2024. [Online]. Available: https://www.netflix.com/
[37] V. K. Butte and S. Butte, “Enterprise Data Strategy: A Decentralized Data Mesh Approach,” in *2022 International Conference on Data Analytics for Business and Industry (ICDABI)*, IEEE, 2022, pp. 62–66. doi: 10.1109/ICDABI56818.2022.10041672.
[38] A. Wider, S. Verma, and A. Akhtar, “Decentralized Data Governance as Part of a Data Mesh Platform: Concepts and Approaches,” 2023, doi: 10.48550/arXiv.2307.02357.
[39] I. Araújo Machado, C. Costa, and M. Y. Santos, “Data-Driven Information Systems: The Data Mesh Paradigm Shift,” *Procedia Comput Sci*, vol. 196, pp. 263–271, 2022, doi: 10.1016/j.procs.2021.12.013.
[40] A. Pakrashi, D. Wallace, B. Mac Namee, D. Greene, and C. Guéret, “CowMesh: a data-mesh architecture to unify dairy industry data for prediction and monitoring,” *Front Artif Intell*, vol. 6, 2023, doi: 10.3389/frai.2023.1209507.
[41] J. Li, S. Cai, L. Wang, M. Li, J. Li, and H. Tu, “A novel design for Data Processing Framework of Park-level Power System with Data Mesh concept,” in *IEEE International Conference on Energy Internet*, IEEE, 2022, pp. 153–158. doi: 10.1109/ICEI57064.2022.00032.
[42] S. Dahdal, F. Poltronieri, M. Tortonesi, C. Stefanelli, and N. Suri, “A Data Mesh Approach for Enabling Data-Centric Applications at the Tactical Edge,” in *International Conference on Military Communications and Information Systems*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICMCIS59922.2023.10253568.
[43] N. Abolhassani, A. Tudor, and S. Paul, “A Data Mesh Adaptable Oil and Gas Ontology Based on Open Subsurface Data Universe (OSDU),” in *International*

Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, IC3K - Proceedings, Science and Technology Publications, Lda, 2023, pp. 29–39. doi: 10.5220/0012160000003598.

[44] N. J. Podlesny, A. V. D. M. Kayem, and C. Meinel, "CoK: A Survey of Privacy Challenges in Relation to Data Meshes," in *Database and Expert Systems Application. Lecture Notes in Computer Science*, vol. 13426, C. Strauss, A. Cuzzocrea, G. Kotsis, A. M. Tjoa, and I. Khalil, Eds., Cham: Springer International Publishing, 2022, pp. 85–102. doi: 10.1007/978-3-031-12423-5_7/COVER.

[45] A. Dolhopolov, A. Castelltort, and A. Laurent, "Exploring the Benefits of Blockchain-Powered Metadata Catalogs in Data Mesh Architecture," Jul. 2023, Accessed: Jan. 30, 2024. [Online]. Available: https://hal.umontpellier.fr/hal-04156089v1

[46] A. Dolhopolov, A. Castelltort, and A. Laurent, "Implementing a Blockchain-Powered Metadata Catalog in Data Mesh Architecture," in *Blockchain and Applications, 5th International Congress. BLOCKCHAIN 2023. Lecture Notes in Networks and Systems*, J. M. Machado, J. Prieto, P. Vieira, H. Peixoto, A. Abelha, D. Arroyo, and L. Vigneri, Eds., Cham: Springer Nature Switzerland, 2023, pp. 348–360. doi: 10.1007/978-3-031-45155-3_35.

[47] T. van Eijk, I. Kumara, D. Di Nucci, D. A. Tamburri, and W.-J. van den Heuvel, "Architectural Design Decisions for Self-Serve Data Platforms in Data Meshes," Feb. 2024, doi: 10.48550/arXiv.2402.04681.

[48] A. Goedegebuure, I. Kumara, S. Driessen, W. Van Den Heuvel, G. Monsieur, and D. Andrew Tamburri, "Data Mesh: a Systematic Gray Literature Review," 2023, doi: 10.48550/arXiv.2304.01062.

[49] Thoughtworks Inc., "Thoughtworks." Accessed: Apr. 22, 2024. [Online]. Available: https://www.thoughtworks.com/

[50] D. Sato and R. Nicholson, "Aligning data architecture with organizational structure in the financial services sector," 2022. Accessed: Apr. 23, 2024. [Online]. Available: https://www.thoughtworks.com/content/dam/thoughtworks/documents/whitepaper/tw_whitepaper_aws_aligning_data_architecture.pdf

[51] INNOQ, "Data Mesh." Accessed: Jan. 30, 2024. [Online]. Available: https://data-ai.innoq.com/en/services/data-mesh

[52] INNOQ, "Regulatory Compliance for Financial Services." Accessed: Apr. 22, 2024. [Online]. Available: https://data-ai.innoq.com/en/use-cases/regulatory-compliance-for-financial-services

[53] PwC International Limited, "PwC: Building trust for today and tomorrow." Accessed: Apr. 22, 2024. [Online]. Available: https://www.pwc.com/gx/en.html

[54] M. Whyte, "Data mesh – The next-generation enterprise data platform?" Accessed: Jan. 25, 2024. [Online]. Available: https://www.pwc.de/en/digitale-transformation/data-mesh-the-next-generation-enterprise-data-platform.html

[55] PwC India, "Financial Services Data and Analytics newsletter," Nov. 2023. Accessed: Jan. 25, 2024. [Online]. Available:

https://www.pwc.in/assets/pdfs/financial-services-data-and-analytics-newsletter-november-2023.pdf
[56] LTIMindtree, “LTIMindtree - Technology Consulting and Digital Solutions Company.” Accessed: Apr. 22, 2024. [Online]. Available: https://www.ltimindtree.com/
[57] D. Verma, “Data Products Using Data Mesh for Different Industries,” 2023. Accessed: Jan. 30, 2024. [Online]. Available: https://www.ltimindtree.com/wp-content/uploads/2023/04/Data-Products-Using-Data-Mesh-for-Different-Industries-POV.pdf
[58] Credera, “Credera.” Accessed: Apr. 22, 2024. [Online]. Available: https://www.credera.com/en-gb
[59] A. Peneda, “Data mesh: How can the insurance sector protect their future?,” Sep. 2021. Accessed: Feb. 02, 2024. [Online]. Available: https://www.credera.com/en-gb/insights/whitepaper-data-mesh-how-can-the-insurance-sector-protect-their-future
[60] Deakin University, “Qualitative study design.” Accessed: Jan. 05, 2024. [Online]. Available: https://deakin.libguides.com/qualitative-study-designs/interviews
[61] Empirical Standards, “Qualitative Surveys (Interview Studies).” Accessed: Jan. 05, 2024. [Online]. Available: https://www2.sigsoft.org/EmpiricalStandards/docs/standards?standard=QualitativeSurveys
[62] M. C. Harrell and M. A. Bradley, “Data Collection Methods: Semi-Structured Interviews and Focus Groups,” RAND Corporation, Dec. 2009. Accessed: Jan. 08, 2024. [Online]. Available: https://www.rand.org/pubs/technical_reports/TR718.html
[63] “Semi-Structured Interviews,” Mar. 2018. Accessed: Jan. 05, 2024. [Online]. Available: http://sociology.fas.harvard.edu/files/sociology/files/interview_strategies.pdf
[64] B. DiCicco-Bloom and B. F. Crabtree, “The qualitative research interview,” *Med Educ*, vol. 40, no. 4, pp. 314–321, Apr. 2006, doi: 10.1111/J.1365-2929.2006.02418.X.
[65] LinkedIn Inc., “LinkedIn.” Accessed: Jan. 06, 2024. [Online]. Available: https://www.linkedin.com/
[66] Zoom Video Communications Inc., “One platform to connect | Zoom.” Accessed: Jan. 07, 2024. [Online]. Available: https://zoom.us/
[67] Otter.ai, “Otter.ai - AI Meeting Note Taker & Real-time AI Transcription.” Accessed: Jan. 14, 2024. [Online]. Available: https://otter.ai/
[68] VERBI GmbH, “MAXQDA | All-In-One Tool for Qualitative Data Analysis.” Accessed: Jan. 14, 2024. [Online]. Available: https://www.maxqda.com/
[69] “Data mesh in the financial sector: Thematic analysis.” Accessed: May 10, 2024. [Online]. Available: https://docs.google.com/spreadsheets/u/1/d/e/2PACX-1vQuESnUiSs79sM5MtLcbdzDe5_QCCPUXZ0gjBrjt443ez--PouATngJMUtl9z88ODsaDtZWv9rbiZAT/pubhtml?gid=0&single=true

[70] C. Richardson, “What are microservices?” Accessed: Apr. 08, 2024. [Online]. Available: https://microservices.io/
[71] Project Management Institute Inc., “Value Streams.” Accessed: Apr. 16, 2024. [Online]. Available: https://www.pmi.org/disciplined-agile/process/value-streams

Appendix

I. Interview Guide

Introductory Questions

1. What is your current role in the financial sector?
2. How long have you been working in this position?
3. What responsibilities do you have in your current role?
4. What has been your overall experience in terms of data management in the financial sector?

RQ Related Questions

RQ1. How is data currently managed by financial service providers?

1. What types of data systems are you most familiar with using in the financial sector?
2. Could you describe a typical data management and structuring process you have experienced in the financial sector?

RQ2. How do the current data management practices limit financial service providers in achieving their business needs and objectives?

1. In your experience, what have been the primary business needs and objectives guiding data management in the financial sector?
2. How have the data management practices you have worked with impacted these needs and objectives?
3. What limitations or challenges have you encountered with the current data management practices in the financial sector?

RQ3. How can the data domains be defined for financial service providers to facilitate achieving business needs and objectives?

1. What would be an effective way to organize data into domains to support business needs and objectives?
2. What are the factors that influence data organization into domains?

RQ4. How valuable could the adoption of a data mesh architecture be for financial service providers?

1. Are you familiar with the concept of data mesh, and if so, what benefits or value could it bring to a financial organization?
2. What would be the main challenges in moving towards adopting a data mesh architecture for a financial service provider?
3. What could be done to address or overcome these challenges?
4. In your opinion, how valuable could the adoption of a data mesh architecture be for a financial service provider?

Concluding Questions

1. Based on your experience, what future trends do you foresee in data management within the financial sector?
2. Before we conclude, would you like to share any additional insights or topics not yet discussed?

II. Schedule of Interviews

Interviewee code	Date	Start time	Duration (hh:mm:ss)
I-01	07.02.2024	15:00	00:57:13
I-02	08.02.2024	16:00	00:58:57
I-03	09.02.2024	14:00	00:57:38
I-04	13.02.2024	10:30	01:11:15
I-05	16.02.2024	12:30	01:28:01
I-06	16.02.2024	15:30	00:43:37
I-07	16.02.2024	17:00	01:02:45
I-08	21.02.2024	11:00	01:10:58
I-09	05.03.2024	13:30	01:02:04
I-10	08.03.2024	16:00	01:12:51

III. License

Non-exclusive licence to reproduce the thesis and make the thesis public

I, **Anette Maria Kuklane**,

1. grant the University of Tartu a free permit (non-exclusive licence) to

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, my thesis

"Data Mesh for Financial Service Providers,"

supervised by Fredrik Milani.

2. I grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in points 1 and 2.

4. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Anette Maria Kuklane

14.05.2024